87-1421

No.____

IN THE SUPREME COURT OF THE UNITED STATES

October Term, 1987

THE CITY OF COLUMBUS, OHIO, et al., Petitioners,

V.

ANN BRUNET, et al., Respondents.

PETITION FOR A WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE SIXTH CIRCUIT

APPENDIX TO PETITION

EILEEN A. GROVES
Assistant City Attorney
90 West Broad Street
Columbus, Ohio 43215
(614) 222- 7385
Counsel of Record

RONALD J. O'BRIEN
City Attorney
Columbus, Ohio
Of Counsel

February 15, 1988

INDEX TO APPENDIX

FILED Aug.25,1987

Nos. 86-3557 and 86-3603
UNITED STATES COURT OF APPEALS
FOR THE SIXTH CIRCUIT

ANN BRUNET, et al.,
Plaintiffs-Appellees;
Cross-Appellants,

v.

On Appeal from the
United States District
Court for the Southern
District of Ohio.

CITY OF COLUMBUS, OHIO, et al.,
Defendants-Appellants;
Cross-Appellees.

_____/

BEFORE: KENNEDY, MILBURN, and NORRIS,
Circuit Judges.

PER CURIAM. Defendants-appellants/cross-appellees the City of Columbus, et al. ("City") and plaintiffs-appellees/cross-appellants Ann Brunet, et al. ("applicants") appeal the District Court's interlocutory judgments of May 11, 1986 and May 30, 1986. Because we find that this Court no longer has jurisdiction to hear this appeal and cross-appeal, we dismiss them.

Plaintiffs Ann Brunet, Lynn Walters, Rebecca Schumacher and Edwina Hornung and the

class they represent challenge the tests that the City has used to hire firefighters since 1975. The named plaintiffs took the tests administered in 1980 and 1984, none of them was selected to be a firefighter. They alleged in a two-count complaint that the 1980 and 1984 tests discriminated against women in violation of Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq. ("Title VII"), and 42 U.S.C. § 1983. They sought injunctive and back pay relief on behalf of themselves and the women they represent.

On May 11, 1986, the United States District court for the Southern District of Ohio entered a judgment for the City on the section 1983 claim. The court also filed separate judgments for the City on the Title VII claim with respect to the 1980 test,and a judgment for the applicants on the Title VII claim with respect to the 1984 test even though these claims

were not made in separate counts nor were the various years' tests separate claims.[1] On June 26, 1986, the District Court entered a judgement nunc pro tunc to May 30, 1986 with respect to the remedy for the Title VII violation found, ordering the parties to comply with the following terms and conditions of the District Court's Opinion and Order of May 30, 1986:

> 1) the City was enjoined from hiring any entry level firefighters on the basis of its new, 1986, test until it had complied with the order;
> 2) the City was required to submit to the Court a report detailing expert analysis of the 1986 test.
> 3) if the Court found the 1986 test to be content valid, then the City was required to report to the Court the results of administering the 1986 test to the incumbent firefighters.
> 4) Once the City had formulated a content valid test and the Court had determined pass/fail scoring procedures, the city was required to administer the new test to all 1984 female applicants who responded to notice.
> 5) Once the City had completed the steps outlined above, they could

> hire on the basis of the results of the 1986 test. However, the City must hire males and females in proportion to the relative proportion of males and females achieving passing scores.
> 6) Since the applicants had prevailed in part upon an issue determining the rights of the parties, they could apply for interim fees.

The City and the applicants each appealed from these judgments. However, since the time of these orders the City has complied with the May 30, 1986 (June 26, 1986 nunc pro tunc order) order and the District Court has approved a new, 1986 test. Opinion and Order of June 12, 1987 at 3.[2]

At the time of these judgments the District Court had not rendered a final decision. Indeed the litigation is still pending before the District Court. A final decision is one which "ends the litigation on the merits and leaves nothing for the court to do but execute the judgment." Catlin v. United

States, 324 U.S. 229, 233 (1945). Since the District Court had not yet determined the legality of the City's new, 1986 test at the time of these appeals, it had not yet ended the litigation on the merits. Thus this Court has jurisdiction to hear these appeals only if they are from "interlocutory orders [s] . . . granting, continuing, refusing or dissolving injunctions or refusing to dissolve or modify injunctions" under 28 U.S.C. § 1292(a)(1), if they fall within the collateral order doctrine, or if they are from judgments on separate counts and the District Court has certified in accordance with Fed. R. Civ. P. 54(b) that there is no just reason for delay. The District Court has not made any such certification. As to section 1292(a)(1), "because § 1292(a)(1) was intended to carve out only a limited exception to the final judgment rule, '[u]nless a litigant can show that an interlocutory order of the district

court might have a serious, perhaps irreparable, consequence, and that the order can be effectually challenged only by immediate appeal, the general congressional policy against piecemeal review will preclude interlocutory appeal.'" Gillis v. United States Dep't of Health & Human Serv's., 759 F.2d 565, 567 (6th Cir. 1985) (quoting Carson v. American Brands, Inc., 450 U.S. 79, 84 (1981)). A collateral order is appealable only if it conclusively determines the disputed question, resolves an important issue completely separate from the merits of the action and is effectively unreviewable on appeal from a final judgment. Stringfellow v. Concerned Neighbors in Action, 107 S. Ct. 1177, 1181-82 (1987).

The judgment of May 11, 1986 that the City's 1984 test violated Title VII is neither an injunction nor a collateral order. The District Court, by its judgment

of May 30, 1987, enjoined the City from hiring on the basis of its new, 1986 test until it had complied with the order of May 30, 1986. That order required the City to develop a new, content-valid test, have the Court approve it, and administer it to the 1984 applicants. Once they had done that, the City could hire on the basis of the 1986 test, but they had to hire males and females in proportion to their presence in the group of candidates who had passed the test. This injunctive order was appealable. However, it is now moot. The City has complied with part of the order by developing a test which the District Court has found to be content-valid and by administering it to the 1984 applicants. At oral argument this Court inquired of the City as to whether there was any part of the order which had not already been performed. The City conceded that the only portion of the order which is not moot is the requirement that they hire a certain

proportion of women. However, the present requirement to hire a certain proportion of women arises from the District Court's Opinion and Order of May 21, 1987. That later provision would require present hiring even if the court were to vacate the injunction of May 30, 1986. Thus the entire May 30th injunction is moot.

The City also appeals that portion of the District Court's Opinion and Order of May 30, 1987 which holds the applicants had prevailed "in part" for the purpose of interim fees. Since the District Court has not in fact awarded interim fees to the applicants, this appeal is premature. See, Myers v. Ace Hardware, Inc., 777 F.2d 1099, 1104-05 (6th Cir. 1985).

The applicants cross-appeal the District Court's finding that the City had not violated Title VII with respect to the 1980 test, the judgment for the City on the section 1983 claim, and subsidiary issues decided by

the District Court in finding that the City had violated Title VII with respect to the 1984 test. Neither the judgment nor the findings meet the requirement for appealability and the applicants conceded this at oral argument.

Arguably the judgment that the City had not violated Title VII with respect to the 1980 test is an order refusing an injunction since the applicants requested relief in the form of an injunction. However, the applicants have not attempted to show that the order could have irreparable consequences, thus section 1292(a)(1) does not provide this Court with jurisdiction with respect to this claim. The collateral order doctrine is not relevant to any of these claims because they all relate directly to the merits of the action. The appellants may not appeal the District Court's findings on issues subsidiary to the judgment on the Title VII claim with respect to the 1984

test. The applicants cannot appeal the district Court's reasoning in reaching the judgment; appellate courts review judgments, not statements in opinions. California v. Rooney, No. 85-1835, slip op. at 3 (U.S. 1987) (per curiam).

Accordingly, the appeal is dismissed as moot. Each party will bear its own costs.

1. The District Court filed an opinion and order discussing these claims on May 13, 1986.

2. In its Opinion and Order of May 21, 1987 the District Court repeated its order to hire males and females in proportion to their relative pass ratio on a revised 1986 test. Opinion and Order of May 21, 1987 at 43.

ISSUED AS MANDATE: November 17, 1987

Costs: None

FILED Oct. 13, 1987

Nos. 86-3557
86-3603

UNITED STATES COURT OF APPEALS
FOR THE SIXTH CIRCUIT

ANN BRUNET, ET AL.,
Plaintiffs-Appellees,
Cross-Appellants,

vs.

ORDER

CITY OF COLUMBUS, OHIO: ET AL.,
Defendants-Appellants,
Cross-Appellees

Upon consideration of the appellants, cross-appellees' petition for rehearing,

IT IS ORDERED that the petition for rehearing be, and it hereby is, DENIED.

ENTERED BY ORDER OF THE COURT

John P. Hehman, Clerk

UNITED STATES DISTRICT COURT
SOUTHERN OHIO DISTRICT
EASTERN DIVISION

ANN BRUNET, et al.,
Plaintiffs,
v. Case No. C-2-84-1973
CITY OF COLUMBUS, et al.,
Defendants.

OPINION AND ORDER

In this action, the named plaintiffs and the class of similarly situated women that they represent challenge certain parts of the tests used by the City of Columbus to select entry-level firefighters since 1979. Plaintiffs Ann Brunet, Lynn Shearrow, Rebecca Schumacher and Edwina Hornung took the tests administered in 1980 and 1984. None of the plaintiffs was selected as a firefighter. Plaintiffs contend in this litigation that they were subjected to discriminatory tests in 1980 and 1984. The defendants are the City of Columbus; the Columbus Civil Service Commission; Dana Rinehart, Mayor of Columbus; and Alphonso Montgomery, Safety Director. For convenience, the defendants

are often referred to as "the City". This action was originally brought under Title VII of the Civil Rights Act of 1964, 29 U.S.C. §s 2000e et seq.; later the complaint was amended to include a claim under 42 U.S.C. § 1983. Plaintiffs seek injunctive and backpay relief on behalf of themselves and the class of women they represent.

Plaintiffs Shearrow, Schumacher and Hornung applied for and took the firefighter selection tests in 1980. Based upon their scores on the exam, plaintiffs were placed on a rank-ordered list of white applicants, to be selected for further consideration in order from that list. Pursuant to this Court's Decree in Dozier v. Chupka, 395 F. Supp. 836 (S.D. Oh. 1975) (Kinneary, J.), the City has maintained dual hiring lists for black and white applicants for firefighter and one-for-one hiring from those lists to remedy past racial discrimination. Of a total of 626 applicants ranked on

the 1980 list, Shearrow ranked 193, Hornung ranked 319, and Schumacher ranked 571. Jt. Ex. 1. Plaintiffs Shearrow and Schumacher timely filed charges of discrimination with the Ohio Civil Rights Commission, Jt. Ex. 30-31, and received right-to-sue letters from the Equal Employment Opportunity Commission. Tr. 214.[1]

Plaintiff Ann Brunet took the entry-level firefighter test held in 1984. She was ranked 464 on the list of non-black applicants. Jt. Ex. 5. Like the other plaintiffs, she was not selected as a firefighter. She timely filed a charge of discrimination and received a right-to-sue letter. Jt. Ex. 32, 26.

In both 1980 and 1984, the firefighter examination consisted of a written examination and a physical test. In 1980, the written test consisted of four sub-tests: a reading comprehension test, a mechanical reasoning test, and two

psychological profiles. Stip. #11. The reading comprehension test was pass/fail; the remaining three tests were scored, and weighted equally to make up 70% of an applicant's total score. Stip. #12. The physical test consisted of seven events, six of which were scored. Timed scores were used to compute a physical exam score which constituted 30% of an applicant's total score. Stip. #15. In 1984, a few changes were made, but the general approach remained the same. The written test consisted of a reading comprehension test and mechanical reasoning test, both of which were scored, and weighted equally to constitute 70% of an applicant's total score. Stip. #31. The physical test was composed of the same events as in 1980 with the exception of one event, which was dropped. As in 1980, the score on the physical test constituted 30% of an applicant's total score. Stip. #35.

In both years, applicants were ranked in order of their total score on separate eligibility lists for white and black applicants. Stip. #23, 36. From time to time, applicants were taken from the lists in order of their rank to be certified to the Columbus Director of Public Safety for consideration for appointment as firefighters. Before being so certified, however, in both 1980 and 1984, applicants were required to pass a ladder test--which involved climbing a ladder to a height of five stories and descending--and a bicycle ergometer test--which measured heart rate in response to physical stress. Stip. #24–26, 37. In addition, applicants were required to pass a medical examination and a background check, and to undergo an interview with a board comprised of members of the Division of Fire. Stip. #27, 37. Applicants who met these requirements were then appointed as firefighters,

as necessary, in the order of the ranking upon the dual lists. Stip. #28, 38. During the life of the 1980 lists, a total of 109 applicants were appointed as firefighters, four of whom were female. Stip. #29. One hundred and twenty-six appointments, including two females, were made from the 1984 list. Stip. #39.

Plaintiffs challenge two components of the firefighter examination: the physical test and the mechanical reasoning test, as discriminatory against female applicants. Plaintiffs contend that the lower scores earned by female applicants on these two components contributed substantially to lower total scores, with the result that fewer female applicants were ultimately selected. Further, they contend, these test components have not been shown by the City to reflect accurately the actual requirements of the job of firefighter.

In their amended complaint, plaintiffs set forth two legal theories. First, they contend that the tests employed by the City have an adverse impact upon female applicants and are not job related. First Amended Complaint, Paragraph 6. This is a theory of prohibited disparate impact under Title VII. Second, plaintiffs contend that the discriminatory acts of the defendants are intentional and violate § 1983. Plaintiffs did not seriously pursue the claim of intentional discrimination at trial or in their post-trial memorandum. In Part I of this Opinion, the Court briefly states its reasons for concluding that plaintiffs have failed to produce sufficient evidence to justify a finding that the defendants engaged in intentional discrimination against women in connection with recruitment of firefighters.

This leaves plaintiffs' adverse impact theory for consideration. In Albemarle Paper Co. v. Moody,

422 U.S. 405 (1975) the Supreme Court described the burdens of the parties in such a disparate impact case as follows:

> In Griggs v. Duke Power Co., 401 U.S. 424 (1971), this Court unanimously held that Title VII forbids the use of employment tests that are discriminatory in effect unless the employer meets "the burden of showing that any given requirement [has] ... a manifest relationship to the employment in question." Id., at 432. This burden arises, of course, only after the complaining party or class has made out a prima facie case of discrimination, i.e., has shown that the tests in question select applicants for hire or promotion in a racial pattern significantly different from that of the pool of applicants. See McDonnell Douglas Corp. v. Green, 411 U.S. 792, 802 (1973). If an employer does then meet the burden of proving that its tests are "job related," it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in "efficient and trustworthy workmanship." Id., at 801.

Id., at 425; accord, Harless v. Duck, 619 F.2d

611, 616 n.6 (6th Cir.), cert. denied, 449 U.S. 872 (1980). The burdens are identical in a case involving alleged discrimination on the basis of sex. Dothard v. Rawlinson, 433 U.S. 321, 329 (1977).

The defendants have argued that the plaintiff class has failed to meet its initial burden of showing adverse impact from either the 1980 or 1984 examinations. Upon consideration of the evidence and the arguments of the parties, the Court concludes, in Part II of this Opinion, that the defendants' arguments are partially meritorious. With respect to the 1980 examination, female applicants who had completed the testing process were selected at essentially the same rate as were similarly situated male applicants. In the judgment of the Court, this fact is fatal to any claim that the 1980 testing and selection process had an adverse impact upon female applicants. However, the

Court further concludes that plaintiffs have carried their initial burden of showing that the 1984 testing and examination process had an adverse impact upon female applicants. As a result of these determinations, only plaintiffs' Title VII claim regarding the 1984 examination remains for consideration.

As a result of plaintiffs' demonstration of adverse impact in the 1984 firefighter examination, it becomes defendants' burden to show that the tests reflect the actual requirements of the job. This burden is often expressed by saying that the defendants must demonstrate that the test is job-related or, equivalently, valid. Having considered carefully the testimony at trial, including the testimony of the parties' respective expert witnesses, and having reviewed the documents submitted as exhibits, the Court concludes, in Part III of this Opinion, that the

defendants have failed to demonstrate the job-relatedness of the 1984 physical examination. On the other hand, the Court further concludes that defendants have adequately justified the mechanical reasoning test, which has also been challenged by the plaintiffs.

In the Court's Opinion, there are two difficulties with the 1984 physical examination. One problem stems from the fact that defendants employ the test scores to rank candidates for selection as firefighters. "Ranking is a valid, job-related selection technique only where the test scores vary directly with job performance." Williams v. Vukovich, 720 F.2d 909, 924 (6th Cir. 1983), citing Guardian's Association of New York v. Civil Service Commission, 630 F.2d at 100. Many more persons apply for the position of firefighter than there are available places. In these circumstances, relatively small differences

in scores can determine whether an individual is selected as a firefighter. If these relatively small differences in test scores reflect likely differences in job performance, then the test is valid, and there is no violation of Title VII. On the other hand, as the Court concludes is the case here, where these differences in scores have not been shown to reflect differences in likely job performance, selection of applicants in accord with such a test is impermissible under Title VII.

In 1975, a report prepared for the City by Battelle concerning hiring criteria for firefighters concluded that physical strength, endurance, agility and health were necessary to perform effectively as a firefighter. Jt. Ex. 24, at 13. The test administered by the City in 1984 is a reasonable test of physical strength in a number of respects that have been shown to reflect the actual physical demands of the job.

It appears also to be a reasonable test of health; at least, no one has raised an issue concerning this aspect of the examination process. However, it is a poor test of endurance, and there is no attempt to test agility. The inevitable result of this narrowed focus upon strength is that relatively small differences in strength will tend to determine whether an individual is selected as a firefighter. There is no guarantee, however, that in selecting stronger individuals, individuals with greater endurance and agility are also being selected. Where a test is used to rank individuals for purposes of hiring, it is important that test cover the range of abilities that are involved in performance of the job. The test administered in 1984 has failed on this count, and is, therefore, invalid when used to rank-order applicants for selection as firefighters.

Having concluded that the defendants have failed to show that the 1984 physical test is job-related, the Court then considers, in Part IV of this Opinion, the remedy to which the plaintiffs are entitled. In light of the absence of substantial evidence of intentional discrimination, the Court concludes that the remedy should be precisely tailored to eliminate the discrimination and restore any individuals to the position they would have occupied but for the discrimination. Accordingly, the Court will order the City to prepare a new physical examination for entry-level firefighters, and to demonstrate its job-relatedness. The City must make the initial decision whether to continue to use a scored physical exam for purposes of ranking, or whether to adopt a pass/fail approach. Whichever approach is adopted, the examination must be approved by the Court before it is administered. Further, before administration of

the new examination, the Court will require the City to provide notice, in a form approved by the Court, of this new examination and the results of this decision to all females who had applied to take the 1984 firefighter examinations. After the new examination has been administered and the results of the examination are before it, the Court will consider retroactive relief and back pay. To the extent that women perform better on the new examination, the Court will presume that they would have so performed on the 1984 examination but for defendants' discrimination. In this circumstance, the Court will fashion a remedy requiring defendants to set aside an appropriate number of places for female applicants in future firefighter classes, and determine the back-pay to be awarded to these applicants. On the other hand, if women as a group perform only as well as, or more poorly than their performance on the 1984 examination, then no retroactive relief would be appropriate.

It is no part of this remedy that the City be required to select women as firefighters in any particular numbers or ratio. Indeed, under Title VII, the gender of an applicant should be irrelevant. As the Supreme Court has explained:

> Nothing in the Act [i.e. Title VII] precludes the use of testing or measuring procedures; obviously they are useful. What Congress has forbidden is giving these devices and mechanisms controlling force unless they are demonstrably a reasonable measure of job performance. Congress has not commanded that the less qualified be preferred over the better qualified simply because of minority origins. Far from disparaging job qualifications as such, Congress has made such qualifications the controlling factor, so that race, religion, nationality, and sex become irrelevant.

Griggs v. Duke Power Co., 401 U.S. 424, 436 (1971). The issue before this Court is not whether women should be firefighters, or how many women should be firefighters. Rather, the issue is

whether the test used by the defendants to select firefighters complies with Title VII. When the defendants administer a valid, job-related examination, that examination will determine how many women are to become firefighters.

I.

Plaintiffs have alleged in their amended complaint that the defendants engaged in intentional discrimination by employing the physical and mechanical reasoning tests to select firefighters and have also addressed this matter in a perfunctory manner in their post-trial memorandum. Plaintiffs contend that intent to discriminate can be inferred from the following evidence. First, prior to 1975, job announcements for the position of firefighter

were restricted to males. Tr. 25. Second, only five of 832 firefighters are women. Tr. 203. Third, plaintiffs have presented evidence about bias against women on the part of the Director of the Training Academy. Tr. 198–202; 819–822. It appears that this led to his removal as head of the Training Academy. Tr. 821. Finally, plaintiffs argue that the defendants, at various times, were aware of less discriminatory testing methods than those they were employing, but refused to adopt them.

However, there is substantial evidence in the record showing that the City made efforts to encourage women to apply as firefighters and to complete the selection process. Marie Hardin, Equal Employment Opportunity Administrator for the City, testified at length about her efforts to recruit females to participate in both the 1980 and 1984 selection processes. Tr. 810–819. These efforts included

maintaining contact with female applicants after their appointment. Tr. 818. Further, although the Court heard testimony from two incumbent female firefighters, Francisca Figueroa and Yolanda Stewart, no evidence of discriminatory treatment was offered by these witnesses. Tr. 156–184; 770–805. In addition, there appears to be no discrimination against women in the administration of the physical examination, as plaintiff Shearrow admitted in her testimony. Tr. 192. Plaintiff Brunet testified that she was permitted to practice the physical examination before taking it and received hints and assistance from firefighters during those practice sessions. Tr. 224–225.

In light of the evidence before it, the Court cannot draw the inference of intentional discrimination suggested by the plaintiffs. Plaintiffs' evidence of intent to discriminate is at best impressionistic. Further, there is substantial evidence

suggesting the absence of discrimination. Accordingly, judgment must be rendered for the defendants on plaintiffs' claim of intentional discrimination under § 1983.

II.

In this section of this Opinion, the Court considers whether plaintiffs have met their initial burden of showing that the examinations administered in 1980 and 1984 had an adverse impact upon the class of women they represent. Having considered the evidence before it and the arguments of the parties, the Court concludes that plaintiffs have failed to show adverse impact in the case of the 1980 examination, but have shown adverse impact in the case of the 1984 examination. Because they present separate questions, each examination will be discussed separately.

Prior to trial, defendants filed a motion

for partial summary judgment, arguing that plaintiffs had failed to carry their initial burden of showing that the 1980 firefighter's examination had an adverse impact upon women. This motion was not ruled upon prior to trial. At trial, defendants renewed their contention at the close of plaintiffs' evidence, seeking dismissal of plaintiffs' claims arising from the 1980 examination. The Court reserved ruling upon defendants' motion and now renders its Opinion.

The facts pertinent to defendants' motion are not in dispute; indeed, they have been stipulated by the parties. In 1980, the Columbus Municipal Civil Service Commission received applications from a total of 1,577 individuals, of whom 83 were females and 1,494 were males. Stip. #9. The Civil Service Commission required all applicants to meet certain minimal requirements, e.g., having completed tenth

grade in school. These requirements eliminated eight male applicants and no female applicants. Stip. #9. Accordingly, 83 female applicants and 1,486 male applicants were invited to the first stage of the 1980 testing process, the written test. Thirty-five female applicants and 387 male applicants failed to appear for the written test. Stip. #10.

In 1980, the written test consisted of four subtests: a reading comprehension test, a mechanical aptitude test, and two psychological tests. The reading comprehension test was graded pass/fail, and applicants who failed were eliminated from further consideration. Three females and seventy-four males failed this test. Stip. #11. The remaining three tests were scored. All applicants who took the written test, including those who failed the reading comprehension subtest, were invited to the next stage,

the physical capabilities test. Of the 48 females invited, 20 failed to appear; 303 of the 1,099 invited males failed to appear. Twenty-eight females completed the physical capabilities test; of these, twenty-five were placed on the 1980 eligibility list. Seven hundred ninety-six males completed the test, and 722 were placed on the eligibility list. Stip. #13. A total of 109 applicants were appointed as firefighters from the 1980 eligibility lists: four were females and 105 were males. Stip. #29. These appointments were made from dual lists for black and white applicants according to a process of one-for-one hiring mandated by this Court's order in Dozier v. Chupka, 395 F. Supp. 836 (S.D. Oh. 1975). Stip. #28. All four female applicants were appointed from the black list.

Defendants argue that, taken as a whole, the 1980 testing process did not have an adverse

impact upon women. Of the twenty-eight females who completed the testing process, four-- or 14% --were ultimately hired. Of the 804 males who similarly completed the process, 105--or 13% --were hired. Thus, defendant asserts, when the process is evaluated from the point of view of its ultimate result, there is no detrimental impact upon women.[2] Defendants' reliance upon hiring ratios among actual applicants appears reasonably grounded in the relevant case law. Berkman v. City of New York, 536 F. Supp. 177, 206 n. 19 (E.D.N.Y. 1982) aff'd, 705 F.2d 584 (2nd Cir. 1983).

In response, plaintiffs argue that the Court should focus upon the components of the testing process, specifically the physical test, and evaluate the discriminatory impact, if any, of these components. Plaintiffs contend that this approach is compelled by the decision of the Supreme Court in Connecticut v. Teal, 457 U.S. 440 (1982). In addition,

plaintiffs offer statistics to show differences in the average scores of men and women on the 1980 firefighter examination. Plaintiffs argue that these differences in mean scores show adverse impact.

With respect to the issue of whether the 1980 firefighter's test as a whole or its components is the appropriate unit of analysis, it is apparent to the Court that the central issue between the parties is the interpretation of Connecticut v. Teal, supra. In Teal, a state agency required that employees achieve a passing score on a written examination in order to be promoted to supervisor. The passing rate on the examination for black candidates was approximately 68% that for white candidates. It was undisputed that the examination, by itself, had an adverse impact upon blacks. Id., at 442, n. 4. However, the score upon the written examination was not the sole criterion for promotion.

Rather, it was used to generate a list of eligible candidates. Selections from the list were made by considering past work performance, recommendations of candidates' supervisors and seniority. The result of this selection process was that approximately 23% of the black candidates on the eligible list were promoted to supervisor, while only 13.5% of the white candidates were promoted. Id., at 444. Thus, the state argued--and this was the sole issue before the Supreme Court--that this "bottom line" result should be considered a complete defense to a race discrimination suit. Even though the state had argued that the bottom line result was a defense, the Court construed the issue as whether plaintiffs had made a prima facie case. Id., at n. 7, and p. 451.

The Supreme Court rejected the "bottom line" approach urged by the state. The Court focused upon § 703 (a)(2) of Title VII, which provides:

> It shall be an unlawful employment practice for an employer to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin.

42 U.S.C. § 2000e-2(a)(2). The Court reasoned that the statute speaks, not in terms of jobs and promotions, but rather "in terms of limitations and classifications that would deprive any individual of employment opportunities." Id., at 448, emphasis in original. Thus, the Court concluded:

> When an employer uses a non-job--related barrier in order to deny a minority or woman applicant employment or promotion, and that barrier as a significant adverse effect on minorities or women, then the applicant has been deprived of an employment opportunity "because of ... race, color, religion, sex, or national origin." --- Relying on § 703(a)(2), Griggs explicitly focused on employment

> "practices, procedures, or tests," 401 U.S. at 430, that deny equal employment "opportunity," id. at 431 ... The examination given to respondents in this case surely constituted such a practice and created such a barrier.

Id., at 448-449.

Teal differs from the instant case in the respect that the challenged component of the selection process, the written examination, was graded pass/fail. Here, however, the challenged portions of the testing process were given a numerical score, which was used, along with other similar scores, to rank candidates on eligibility lists. Thus, the written examination in Teal constituted a "barrier" in the sense that it precluded candidates from further consideration. The challenged components of the testing process here, even though lower scores on these components may lessen a candidate's overall chance of acceptance, do not preclude further consider-

ation of that candidate. The question that the Court must decide is whether this difference amounts to a distinction.

For the following reasons, the Court concludes that Teal is distinguishable from the instant case and, therefore, rejects plaintiffs' contention that the bottom line result does not negate adverse impact. In Teal, the actual holding of the Court is:

> [R]espondent's claim of disparate impact from the examination, a pass-fail barrier to employment opportunity, states a prima facie case of employment discrimination under § 703(a)(2), despite their employer's nondiscriminatory "bottom line," and that "bottom line" is no defense to this prima facie case under § 703(h).

Id., at 452. Thus, the holding is limited by its terms to a pass/fail barrier. Concededly, there is language in the opinion that sweeps more broadly. It does not appear that this language is essential to the reasoning of the majority opinion, however. The critical premise

in the majority's reasoning is that the pass/fail subtest eliminated individuals from further consideration.

In addressing the precise issue before the Court, Schlei and Grossman, in their widely cited text on employment discrimination, comment:

> It seems probably that Teal's rejection of the bottom line approach with respect to components that constitute a "pass/fail barrier" to further consideration in the selection process will not be applied to multicomponent selection processes where all candidates complete all components of the process before the selection is made. Although the majority did not specifically address this issue, the Second Circuit decision below, which was affirmed, specifically so held, and the four Justices in dissent so interpreted the majority opinion.

B. Schlei & P. Grossman, Employment Discrimination Law (2nd ed. 1983), at 1377-1378. In Teal, the Second Circuit had written:

> Where all of the candidates participate in the entire selection process,

> and the overall results reveal no significant disparity of impact, scrutinizing individual questions or individual sub-tests would, indeed, "conflict[] with the dictates of common sense."

Teal v. State of Connecticut, 645 F.2d 133, 138 (2nd Cir. 1981), aff'd, 457 U.S. 440 (1982), quoting Kirkland v. New York State Dept. of Correctional Services, 374 F. Supp. at 1370.

In Smith v. Troyan, 520 F.2d 492 (6th Cir. 1975), cert. denied, 426 U.S. 934 (1976), the Sixth Circuit held that, where the overall examination process had no disparate racial impact, it was error to require a defendant to prove that a component of the overall process was job-related, even though blacks fared less well on that sub-test. As in the instant case, the score on challenged subtest was added to scores on other subtests and used to rank eligible candidates. 363 F. Supp. at 1134-1135, 1144-1145. In these circumstances, the Court concluded, the plaintiff had

failed to demonstrate prima facie that the test was unlawfully discriminatory. Id. at 497. The Court observed:

> though general ability, or intelligence, tests have often been invalidated for their racially disproportionate impacts ... (cites omitted) ..., the disproportionate impacts have been in the hiring, rather than in the test results in and of themselves.

Id., at 497-498. Teal does not squarely overrule this result, which must, therefore, be considered to be controlling law in this Circuit.

The Uniform Guidelines on Employee Selection Procedures ("Guidelines"), 29 C.F.R. §s 1607.1 et seq., also support the view that individual components of a testing procedure need not be justified by an employer where the entire testing procedure does not have an adverse impact. Where the total selection process does not have any adverse impact,

> [t]he Federal enforcement agencies ... will not expect a user to evaluate the individual components for adverse

> impact, or to validate such individual components, and will not take enforcement action based upon adverse impact of any component of that process, including the separate parts of a multipart selection procedure ...

29 C.F.R. § 1607.4(C). Although not binding upon this Court, the Uniform Guidelines are entitled to substantial deference as the interpretation of the Act by the enforcing agency. Albemarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975).

Furthermore, there is some question whether Teal should be applied in case, like the instant case, which is brought as a class action. In Coser v. Moore, 587 F. Supp. 572 (E.D. N.Y. 1983), aff'd, 739 F.2d 746 (2nd Cir. 1984), the district court construed Teal to be inapplicable in the case of a class action by women alleging system-wide discrimination on the basis of sex. The court interpreted Teal as involving a claim by

individuals who failed a written examination with a proven adverse impact.

> The error of the district court in Teal was to foreclose proven and unrebutted individual claims of discrimination by looking to an employer's treatment of a group.

Id. at 588, emphasis in original. The case before the Coser court involved an attempt by a class of women to prove sex discrimination in hiring and promotions on a university-wide basis. To prove their case, the plaintiff class presented evidence of under-utilization of women in specific departments and divisions of the university. In response, the university presented evidence of lack of discrimination in university-wide hiring. The plaintiffs argued that this evidence was no defense under Teal. The issue was thus analogous to the issue presented by the instant case.

The court rejected this reliance on Teal, reasoning as follows:

> Unlike the individual plaintiffs in Teal, plaintiffs here are a class of

> women seeking to prove by disparate impact analysis that Stony Brook has a pattern and practice of discrimination against women. If successful, that finding would then enable individual plaintiffs to rely on an inference of discrimination when they seek to prove their individual claims ... [T]he issue is whether Stony Brook's neutral criteria have an adverse impact upon a group, and upholding Stony Brook's defense against plaintiffs' class action claims would not foreclose valid individual claims of discrimination, as the "bottom line" defense did in Teal.

Id. at 588, emphasis in original. In the instant case as in Coser, plaintiffs are asserting a group claim. It follows that their proof of adverse impact necessarily depends upon the fortunes of the group. Thus, Teal's focus upon the individual appears misplaced in the context of the instant case.

Alternatively, plaintiffs propose to demonstrate adverse impact of the entire examination and selection process by focusing upon the differences in average scores of men and women on the 1980

firefighter exam. According to calculations made by plaintiffs' expert Dr. Joseph Cranny, the average (or mean) total score of females on the 1980 examination was 80.13, while the average score for males was 85.21. Jt. Ex. 6. On the physical agility test alone, females averaged 36.00 while males averaged 49.98. Cranny also calculated a correlation of .36 between the score on the physical test and the overall test score. This showed, in his words, that there is a "slight tendency" for people who do well on the physical test to do well on the total test. Cranny Depo. of Dec. 21, 1984, at 27; Tr. 260. Finally, Cranny calculated that the statistical likelihood of these differences in scores arising by chance was extremely small.

Plaintiffs argue that the difference in average scores means that women have less chance of being selected as firefighters than men. This

lessened opportunity arises because candidates are selected in order of their scores upon the tests. Furthermore, a significant part of these differences in scores arise from the physical tests challenged in this litigation. Accordingly, plaintiffs conclude that women have been denied an equal opportunity to be considered for the position of firefighter.

Upon consideration, the Court declines to draw the inference of denial of equal opportunity from the differences in average scores. It is surely relevant to note that the actual result of the 1980 selection process was that women were hired at a slightly higher rate than men. It is difficult to ascribe any meaning to the notion of denial of equal opportunity when it is considered in light of this fact. Title VII does not require employers to equalize the probabilities of hiring of the average members of two groups. Rather,

it requires that actual individuals enjoy opportunities for employment free from discriminatory barriers.

The reliance upon differences in mean scores is misplaced for an additional reason. There are far more applicants than there are available jobs in the Columbus Fire Department. Consequently, only the applicants earning the highest scores have any realistic chance of being hired. Thus, it is the impact of the examination upon the highest scorers, not the average impact that is significant. See United States v. City of Chicago, 549 F.2d 415, 429 (7th Cir.), cert. denied, 434 U.S. 875 (1977). Plaintiffs' statistical expert admitted at trial that it was possible that there be significant differences in average scores for men and women on a test and yet that selection ratios be essentially the same due to the fact that all selections would occur from only a small region of the

distributions. Tr. 376-377.

Plaintiffs contend that the use of mean-difference analysis to show adverse impact was approved by Judge Duncan in Police Officers for Equal Rights v. City of Columbus, No. C-2-78-394 slip op. (S.D. Oh. 1985). One issue in that case was whether the sergeants promotional examinations administered by the Columbus Police Department had an adverse impact upon black police officers. Dr. Joseph Cranny appeared as an expert witness for the plaintiffs. He sought to show adverse impact by three methods: examination of selection ratios under the 4/5's rule of the Guidelines, mean difference analysis, and analysis of pass/fail ratios. The Court concluded that plaintiffs had proven adverse impact under the 4/5's rule. Id. at 88. The Court also noted that plaintiffs had shown a difference in mean scores. Id. at 89. Thus, the case cannot properly be relied upon to support

the contention that mean-difference analysis alone can be relied upon to prove adverse impact. The same is true of Walls v. Mississippi State Dept. of Public Welfare, 542 F. Supp. 281, 293 (N.D. Miss. 1982), aff'd in relevant part, 730 F.2d 36 (5th Cir. 1984) and Thomas v. City of Evanston, 610 F. Supp. 442, 427 (N.D. Ill. 1985), both of which are also cited by plaintiffs.

One case that does support the plaintiffs' reliance on differences in average scores is Burney v. City of Pautucket, 559 F. Supp. 1089 (D. R.I. 1983). One issue in the case was whether physical agility requirements of a police academy had an adverse impact upon women. In order to graduate from the police academy, a recruit was required to score at least a "C" in each course, including a physical test. The score in the physical test was based equally upon performance upon certain physical tests and the

subjective estimate by instructors of the recruit's achievement and attitude. Id., at 1095-1096. Women earned lower scores on the test than did men. The defendants argued, however, that, notwithstanding their lower scores on the physical tests, all of the women who had entered the academy had graduated. Further, their scores on the physical test did not prevent women from graduating at or near the top of their classes. Id., at 1099. The Court rejected these argument, citing Teal for the proposition that such "bottom-line" arguments were no defense.

This Court is unpersuaded by the reasoning of the Burney court. The plaintiff in Burney had been dismissed from the police academy for accumulating excessive demerits in the physical training program. Id., at 1100. Thus, as to her, adverse impact was established by the fact of her dismissal. The average

scores of women on the physical tests are irrelevant to this. More generally, the Burney court, perhaps because it was faced with a case involving an individual claim, appears to have confused the theories of disparate impact and disparate treatment. In any event, it appears that Burney is out of line with the great weight of authority.

Plaintiffs also assert that hiring ratios are unreliable in the instant case due to existence of dual hiring lists for black and white firefighters. The four women hired in 1980 were all selected from the black list. The list of black candidates was substantially shorter than the white list, and thus the process of one-for-one hiring led to hiring from further down the black list. Had there been only one list in 1980, plaintiffs contend, no women would have been hired. Thus, but for the dual lists, no women would have been

hired from the 1980 lists.

This argument is beside the point, even though it may well be factually correct. It is beside the point because the narrow issue presently before the Court is whether plaintiffs, as representatives of a class, have proven adverse impact by the 1980 firefighter examination. This is plaintiffs' initial burden, and must be carried before defendants are required to justify the examination by showing that it is job-related. Whether plaintiffs would more easily have been able to carry their burden had things been different in 1980 is irrelevant. If the class of female applicants in 1980 was not adversely affected by the firefighter examination, then defendants are not liable for their acts connected with the 1980 exam and plaintiff are not entitled to a remedy with respect to that exam. The Court must decide a case such as instant one upon

the facts before it, not upon theoretical possibilities. See Schlei and Grossman, supra, at 102 n. 94.

In his testimony at trial, plaintiffs' expert witness suggested that the equivalence of the hiring ratios in 1980 for male and female applicants was a "complete statistical artifact." Tr. 282. This artifact arose because selection ratios for both males and females derive from large numbers of applicants and small numbers of appointments. Tr. 281–282. Even if this is correct, it is of no consequence for this case. It is the plaintiffs' burden to prove adverse impact, not the burden of the defendants to prove absence of adverse impact.

The Court also declines to assign any significance to the fact that 20 of 48 --or 42%--female applicants failed to appear from the physical exam, while only 303 of 1099 males--or

about 28%- failed to appear. It is true that courts must be mindful of the possibility of deterrence of applicants before relying upon data regarding actual applicants. Dothard v. Rawlinson, 433 U.S. 321, 330 (1977). While these numbers might suggest that some female applicants are deterred from appearing for the physical examination, compare Tr. 286-287 and Jt. Ex. 8, this suggestion is not supported by the evidence produced at trial. At trial, Dr. Gerald Barrett testified that he had made an informal survey of fire testing dropouts in the City of Akron. He found that women and blacks tend to drop out of the testing process at a higher rate than white males. He attributed this to a variety of factors, including change of career orientation and increased knowledge about the job of firefighter. Tr. 663-664. This testimony was corrobrated by the testimony of Marie Hardin based upon her experiences in Columbus, Tr. 817,

as well as the statement of named plaintiff Hornung that she is no longer interested in becoming a firefighter. Tr. 188.

For these reasons, the Court concludes that plaintiffs have failed to prove adverse impact from the 1980 firefighter's examination. Accordingly, defendants' motion to dismiss plaintiffs' Title VII claims regarding the 1980 examination must be GRANTED. Fed. R. Civ. P. 41(b).

As in the case of the 1980 examination, the facts relevant to adverse impact in the 1984 examination have largely been stipulated. In 1984, a total of 2,886 males and 354 females appeared for the written test. Stip. #32. Four hundred and fifteen males and fifty-two females failed the written test. Consequently, 2,471 males and 302 females were invited to take the physical test. Stip. #33. Of those invited, 1,343 males and 83 females appeared and

completed the physical test. Stip. #35. Two females and 124 males have been selected from the 1984 eligibility lists; no further selections from the 1984 list are anticipated. Stip. #39.

In 1984, the selection ratio for women was two out of 83, or 2%; for men, it was 124 out of 1,343, or 9%. The Guidelines have suggested as a rule of thumb that if the selection ratio of the protected group is less than 80% of the selection ratio of the non-protected group, there is likely to be adverse impact in the selection process. 29 C.F.R. § 1607.4(D). Here, the selection rate for female applicants is only about 22% that for male applicants. Further, Dr. Cranny testified at trial that he had performed a chi-square analysis upon these selection ratios, to determine the probability that these observed differences in selection ratios arose by chance. He testified that,

using a one-tailed test, the observed difference was significant at the .05 level, that is, that there is only one chance in twenty that it was the mere result of chance. Tr. 285. Dr. Cranny admitted that the chi-square test was not significant if a two-tailed test was employed. Id. Although defendants question this use of a one-tailed test, the Court concludes that it is appropriate where, as here, the raw numbers indicate that women are selected at a lesser rate than men. In these circumstances, the question being asked is whether this apparent difference is real or a statistical artifact. This question is appropriately answered by a one-tailed test. There is no indication in this record that in reality women are being selected a higher rate than men in 1984.

The Court concludes from this showing of violation of the 80% rule and the chi-square analysis that there was adverse impact upon

women in the 1984 firefighter examination taken as a whole. This conclusion is corroborated by plaintiffs' evidence regarding differences in mean scores of men and women upon the exam. Tr. 266-271; Jt. Ex. 7. As explained above, differences in mean scores may properly be relied upon to corroborate a showing of adverse impact by the 80% rule or chi-square analysis. Police Officers for Equal Rights v. City of Columbus, supra, at 88-89. Thus, plaintiffs have met their initial burden with respect to the 1984 firefighter examination.

III.

Because plaintiffs have shown adverse impact upon women in the 1984 examination, it becomes the defendants' burden to show that the test has a "manifest relationship to the employment in question." Griggs v. Duke Power Co., 401 U.S. 424, 432 (1971). In making this showing, "[t]he touchstone is

business necessity." Id., at 431. The standard of proof of job-relatedness has been stated by the Supreme Court as follows:

> [D]iscriminatory tests are impermissible unless shown, by professionally acceptable methods, to be "predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated." 29 CFR § 1607.4(c).

Albermarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975). In this section of this Opinion, the Courts makes its findings of fact and conclusions of law in support of its determination that defendants have failed to carry their burden.

The 1984 Examination

As has been noted, the 1984 firefighter examination consisted of a written test and a physical test. The written test had two components,

a reading comprehension test and a mechanical reasoning test. Plaintiffs' expert, Dr. Cranny, performed a statistical analysis of the scores of men and women on the 1984 examination and its various components. This analysis is not challenged by the defendants. On the total examination, men, as a group, achieved an average score of 78.5, while the average score of women, as a group, was 65.0. Jt. Ex. 7, at 5. Womens' scores ranged from about 46 to about 82.4[3] while mens' scores ranged from 0 to about 93. Jt. Ex. 7, at 5; Tr. 267-269. These total scores were the result of a series of statistical manipulations to standardize the raw scores on the various component tests. Stip. #36. No issue has been raised regarding the propriety of these statistical manipulations.

The highest score earned by a woman on the 1984 examination was 82.4. Three hundred and

fifty-five white males and twenty-one black males earned higher scores. A total of 126 individuals were ultimately hired as firefighters from the 1984 lists. Had there not been dual hiring lists mandated by Court order, consequently, no females would have been hired as firefighters. In fact, two females were hired, both from the black eligibility list.

The differences in male and female total scores resulted primarily from lower female scores on two components of the total test: the physical test and the mechanical reasoning test. Plaintiffs' challenge is directed to these two components. There was no significant difference between the sexes on the reading comprehension test. Jt. Ex. 7, at 5; Tr. 270. On the mechanical reasoning test, men earned an average score of 19.6, while womens' scores averaged 15.1. Jt. Ex. 7, at 5. The greatest disparity occurred on the physical test, where men averaged 76.1,

while women averaged 44.5. A statistical analysis of test scores by defendants' expert, Dr. Frank Landy, confirmed what is apparent from the raw numbers: the differences in male and female total scores are due primarily to the differences in scores upon the mechanical and physical tests. Jt. Ex. 11; Tr. 275-278.

The 1984 firefighter physical capability test consisted of seven events. All seven events had pass levels, and it was necessary that a candidate pass in order to be considered for hiring. However, the pass levels were set very low, and it appears very few persons failed the physical exam. Five of the six events were scored. Jt. Ex. 52, 53. The events were administered to groups of applicants, at approximately ten minute intervals.

1) Beam Walk: Applicants were required to walk the length of a twenty foot beam that

was four inches wide while carrying a roll of hose. The event was pass/fail only, and was not timed. Three tries were allowed. In 1984, one male failed the beam walk. Stip. #36. Because virtually everyone passed this event, it is of the little consequence in this litigation.

2) Manual Dexterity: Applicants were required to screw three metal plugs into three threaded intakes on a piece of fire equipment, a multiversal, and then unscrew them. The event was timed, and a higher score was earned by completing the event more quickly. At a minimum, the event had to be completed in two minutes. In 1984, women scored about the same on this test event than did men. The average time of women was 26.8 seconds, while that of men was 25.5 seconds. Jt. Ex. 7, at 5; Tr. 272. This difference, however, was not statistically significant. Tr. 272.

This event had been recommended for

inclusion on the test by the Five Training Academy staff, who had experienced problems with recruits who lacked manual dexterity. The suggestion of plaintiffs' expert, Dr. John Magel, an exercise physiologist, that women might be disadvantaged on this event due to less experience with tools than men, Tr. 467-468, is contradicted by the essential equality of average scores. See also Tr. 735-736. This event is a direct simulation of a common firefighting task, as firefighter Yolanda Stewart testified. Tr. 779-780. Francisca Figueroa suggested that there could be a problem with failing to line up the threads properly if one tried to work too fast on the job. Tr. 164-165. This does not appear to be a serious problem, however.

3) Sandbag Drag or Carry: Applicants were required to carry or drag a sand dummy through a designated serpentine course defined by a line

on the floor running around a number of poles. The dummy was the approximate size of a small duffle bag, with straps to grip, and weighed 125 pounds. The event was timed. Further, if the applicant chose to drag the dummy, or dropped it on any part of the course, the time was doubled. Also, if a pole was knocked over, a two-second penalty was imposed.

Men performed substantially better than women on this test event. The mean time for men was 19.5 seconds, while that of women was 38.2 seconds. Because the event was timed, the lower score is better. This difference is statistically significant. Jt. Ex. 7, at 5; Tr. 272-273.

The event was designed to test an applicant's ability to drag or carry adults or children. Jt. Ex. 23. The event is an imperfect simulation. It appears from both expert and firefighter testimony that the weight

of the bag is reasonable. Tr. 718–723; 784. However, the shape of the bag makes it awkward to carry, depriving individuals of the opportunity to use lifting techniques and leverage. Tr. 468–471; 786–787. An articulated dummy could readily have been used. Tr. 324. There is little sense to be made of doubling the score if the bag was dragged; the testimony at trial was that victims are typically dragged from a building, due to the presence of smoke. Tr. 69; 169; 827–828. No rationale appears for placing a premium on extreme speed; the testimony at trial indicates that the speed necessary depends on the circumstances. Tr. 108–169; 827. This event measures primarily upper body strength and anaerobic capacity.

4) Pike Pole Pull: The applicant pulls on a handle attached to a rope which runs through pulleys and is attached to a 75 pound weight. A repetition consists of pulling down the handle until it

strikes a stand, and then returning the handle back to its original position; this, of course, involves lifting and lowering the 75 pound weight the distance of travel of the handle. The entire event lasts one minute. To pass, five repetitions must be completed in that time. The event is also scored: the more repetitions completed, the higher the score.

The scores of males were substantially better than those of females. On the average, men accomplished 58.2 repetitions, while women performed 38.9 repetitions. Jt. Ex. 7, at 5. Unlike the other timed events, here the higher raw score is better. Tr. 273-274.

The pike pole pull is a rough simulation of the actual use of pike poles, a rod with a hook on the end, to rip out walls and ceilings to search for fires. From the testimony at trial, the Court concludes that the seventy-five pound weight reasonably reflects

the physical demands of the job. Jt. Ex. 21; Tr. 708-717. However, the simulation of the job is questionable, in a number of respects. Actual use of a pike pole involves both push and pull phases; the event tests only the pull phase. No rationale appears for the requirement of hitting the stand with the pole; this appears to be merely a device for score-keeping without any analogue in the actual use of the pike pole. It also appears that shorter persons--women tend to be shorter than men--were slightly disadvantaged by the event, because they could use their entire body to less advantage. Tr. 475; 715-716. This bias could have been eliminated by making the apparatus adjustable. And, the test appears to over-emphasize speed as compared with actual practice. The experts agree that the event measures upper body strength and the anaerobic capacity of the upper body. Jt. Ex. 20, at 3-4; Tr. 474.

5) Equipment Hoist: The applicant pulls a rope that runs over a roller to lift a sixty-five pound weight to a third-story window. The event is timed, and the more quickly the weight is raised, the better the score. After it is raised, the weight must be gently lowered to the ground. The lowering is not timed; however, if the weight is dropped, a penalty is assessed. In this timed event, the mean score for men was 10.7 seconds; the mean score for women was 26.9 seconds. Jt. Ex. 7, at 5.

The event was designed to simulate raising ladders and hose by means of a hose roller, an actual piece of fire equipment. Jt. Ex. 23. It does not appear, however, that hoisting is done very often in actual firefighting, mainly because the roller takes too long to set up. Tr. 160-161, 175-176, 778. There is no

indication why the weight of sixty-five pounds was chosen. Taller persons appear to enjoy a slight advantage in the event. Tr. 729. The experts agree that the test measures primarily muscular strength and anaerobic capacity of the arms. Jt. Ex. 20, at 4; Tr. 476-477.

6) Stairway Climb: The applicant was required to climb six flights of stairs and descend as rapidly as possible while wearing fire gear and carrying equipment, a roll of hose. The fire gear and equipment weighed about forty-seven pounds. The event was timed, and the score depended upon how quickly the event could be completed.

The mean score for men was 65.7 seconds, while that of women was 102.2 seconds. Jt. Ex. 7, at 5. The standard deviation of the men's score was 12.6; this means that approximately two-thirds of the male applicants in 1984 completed the event in a time ranging between 53.1 seconds and 78.3 seconds.

The standard deviation of the womens' score was 22.6, so that the comparable range was 79.6 to 124.8 seconds.

Firefighters must frequently climb stairs, though it appears that in tall buildings they use elevators when possible. Sometimes firefighters must climb six or more stories. Tr. 763; 788–789. When equipment must be carried up many stories, it is shuttled up two or three flights of stairs at a time in a relay; this operation, called staging, is more efficient. Tr. 887. Firefighters infrequently run up stairs, both for safety reasons and to marshall their energy to perform when they arrive at the fire. Tr,. 162–164; 789; 843–844. This test measures anaerobic power to sprint; performance does not depend primarily upon cardiovascular endurance or aerobic capacity. Tr. 505–510, 512; 547; 766–767.

Three events on the test--the sandbag drag

or carry, the pike pole pull, and the equipment hoist--measure primarily upper body strength. Further, the test tends to measure anaerobic capacity of the various muscles used. No event measures primarily aerobic capacity. The ten minute resting period between events contributed to the overall anaerobic character of the test. This observation that the various test events tend to measure similar physical abilities is confirmed by the statistical analysis performed by defendants' expert, Dr. Landy. This shows quite substantial statistical correlations among scores on the various events. Jt. Ex. 11.

Two of these events--the beam walk and manual dexterity test-- had no significant impact upon the relative scores of men and women. Thus, the issues in this case turn upon the job-relatedness of four test events: the sandbag drag or carry, pike pole pull, equipment hoist, and stairway climb. These events

were timed, with the exception of the pike pole pull, where the number of repetitions determined the score. In all cases, speed was of the essence. On the three timed events, the average time of women was roughly twice that of men. In the pike pole pull, men completed approximately 50% more repetitions, on the average. These differences determined the differences in total score between men and women upon the physical portion of the examination. And, to a substantial extent, they determined the differences between men and women in total score upon the written and physical examinations. Accordingly, the fairness of 1984 physical test stands or falls upon the validity of these four test events.

There is a very scanty record regarding the other component of the 1984 firefighter examination under challenge here, the mechanical reasoning test. In 1984, it consisted of thirty written

questions, and was scored by adding the number of questions answered correctly. Stip. #31. It constituted 35% of an applicant's total score. Id. There is no indication in the record why this 35% weighting was selected. Nor is there any indication of the content of the mechanical reasoning test, other than that conveyed by its name.

As noted above, women scored less well than men on the mechanical test, averaging 15.1 as against 19.6, respectively. Jt. Ex. 7, at 5. The range of men's scores from 9 to 29 was somewhat higher than women's score range, which was from 8 to 23. Scores upon the mechanical reasoning test were highly correlated with total test scores; the correlation coefficient was .85. Jt. Ex. 11.

Test Development

An event that looms large over the test

development process is this Court's judgment and decree in Dozier v. Chupka, 395 F. Supp. 836 (S.D. Oh. 1975). In Dozier, this Court concluded that the division of fire had employed standards and criteria for the selection of firefighters that had a racially discriminatory impact upon members of the plaintiff class, black male applicants for the position of firefighter. In 1973, the fire department had administered a written aptitude test to applicants. The Court considered the validation attempts undertaken by the defendants and concluded that the written examination had not been validated. 395 F. Supp. at 854. On April 16, 1975, the Court entered a remedial decree enjoining the defendants from further discrimination on the basis of race. Further, the Court ordered the defendants to develop criteria for selection of firefighters and to validate these criteria in compliance with the Equal Employment

Opportunity Guidelines on Testing as set forth in 26 C.F.R. §s 1607.1 et seq. Id. at 859-860.

In 1973, prior to the entry of the Dozier decree, a two-step selection process had been used by the fire department. First, all applicants took a written examination; to be considered further, an applicant must pass that examination. The next step was a physical agility test, which also was initially graded pass/fail. Applicants who passed both tests were then ranked on an eligibility list, their relative position being determined by adding their two test scores and certain bonus points, if any, for military service. Then, a background investigation were conducted. 395 F. Supp. at 840-841. Candidates who were not removed from the eligibility list on the basis of the background investigation were appointed on the basis of total scores. Id., 844; Jt. Ex. 16, at 1-4.

With some modifications, this same general approach to firefighter selection, i.e., ranking applicants according to scores on written and physical exams, was used in the 1980 and 1984 firefighter selection process.

At the time the remedial decree was entered in Dozier, the City had hired Battelle to conduct an analysis of the job of firefighter. The City had informed the Court at the time of the Dozier decree of its intention of doing so, and the Court noted this fact. 395 F. Supp. at 859. Battelle submitted a document entitled "Final Report of Hiring Selection Criteria for the Entry-Level Firefighter" to the Civil Service Commission on June 20, 1975. Jt. Ex. 24. The report proposed hiring selection criteria pertaining to physical abilities, sensory abilities, communication skills, reasoning and judgment skills, and personal and interpersonal characteristics. Id., at 15-22.

The report included an assessment of the physical demands of job. This involved weighing equipment, determining hose recoil pressures, and measuring the size of windows through which firefighters must sometimes crawl. The Battelle study did not propose tests for selecting firefighters; rather, it set forth criteria that tests should be developed to measure. The study concluded that strength, endurance and agility were the most important physical characteristics of firefighters. Id., at 9.

Testing of applicants for firefighter was conducted by the Civil Service Commission in 1975 and 1978. Stip. #1. These were the first tests that were open to female applicants. Prior to 1975 the job announcement for firefighters restricted applicants to males. Tr. 25. In both 1975 and 1978, the Civil Service Commission used a written reading test, a physical agility test, and a battery of tests selected

by Dr. Gerald Barrett to determine mechanical comprehension, math ability and certain personality characteristics. Jt. Ex. 16, at 1–7. The physical agility examinations in 1975 and 1978 were graded pass/fail. Tr. 26; Jt. Ex. 48. Candidates were chosen from the dual hiring lists ordered in Dozier vs. Chupka in order of their written test scores. Stip. 4–5. In 1978 two females were appointed as firefighters, the first females to be so appointed. Stip. #7.

At trial, the Court received into evidence certain documents pertaining to the 1975 firefighter examination. The 1975 test consisted of the following events: bent-knee sit-ups; ladder climb; driver capability (to determine if person is of a size to drive a fire truck); weight lift and twist; stairway climb; ladder raise; push-ups; beam walk with hose; and dummy dodge run. Jt. Ex. 48. It does not

appear that the job analysis performed by Battelle played any role in developing these test events. The Battelle study had recommended that the physical test should duplicate actual physical activities performed by firefighters. Jt. Ex. 24, at 23, which is not apparent in this test. Further, the test scoring procedures are dated June 1975 and appear to derive largely from recommendations made by the Bureau of Training of the Division of Fire in March of 1975. Jt. Ex. 47. The Battelle report was dated June 20, 1975. There is no evidence in the present record about the content of the 1978 examination.

In late 1978, the City began to develop a new physical test for firefighter. Tr. 28. The impetus for this development was the decision of Judge Duncan in Brandt v. City of Columbus, Case No. C-2-75-425 (S.D. Oh. Oct. 5, 1978), a class action alleging sex discrimination in the Columbus Police

Department. In that decision Judge Duncan concluded that the physical agility test used by the police department for selection of recruits failed to meet the validity standards set forth in the Uniform Guidelines, and, therefore, was unlawful under Title VII. Because of the similarities between the testing procedures struck down in Brandt and those used by the fire department, the City reexamined the test procedures. Tr. 29.

In November 1978, Dr. S. David Kriska, who is in charge of personnel testing for the Civil Service Commission, drafted a memorandum reviewing the adequacy of the current firefighter physical examination in light of Brandt. Jt. Ex. 22. Kriska examined the various test events in light of the Battelle job analysis; this appears, from the record, to be the first time that this was done. He recommended that three test events--the sandbag lift and carry, push-up and sit-up events--be eliminated for

lack of job-relatedness. Also, Kriska proposed new events to test upper body strength and endurance, both of which were found relevant by the Battelle study. Plaintiffs' Ex. 3, at 2; Tr. 40. Kriska also noted that setting of passing scores was likely to be a problem, since there was likely to be an adverse impact upon women. Kriska proposed that a modified test should be given to a random sample of firefighters of various ages as well as a group of women likely to be representative of the probable applicant population to determine pass points. Id., at 4; Tr. 34-35. However, neither the 1980 nor 1984 physical tests has been administered to incumbent firefighters. Tr. 43-45. Indeed, it is unclear whether any of the physical tests from 1975 forward has ever been administered to incumbent firefighters.

In May 1979, Julia Ingram, an employee of the Civil Service Commission, and David Kriska issued a report proposing a physical test for firefighter. Jt. Ex. 49. In preparing the report, they had consulted with Dr. Edward Fox, an exercise physiologist and expert witness in Brandt. Id., at 2; Tr. 28. The proposed test included seven events: a beam walk with hose; ladder climb; ladder draw and carry; hose drag; blind hose follow (crawling in fire gear wearing an opaque face mask, following a hose through a predetermined course); stairway climb; and bicycling. Jt. Ex. 49, at 1-4. Each test event was to be graded pass/fail; failing any event would eliminate the candidate from further consideration. This 1979 proposal represents the culmination of efforts to rethink physical testing of firefighters in light of the Brandt decision.

However, before these 1979 suggestions were

were accepted, a critical change in thinking and approach occurred. In 1980, a job analysis was performed by Ingram for the Civil Service Commission. Jt. Ex. 18. Based upon this job analysis, Ingram and Kriska concluded that the work of firefighting was largely physical, and that better firefighters were distinguished by the ability to excel while performing physical tasks. Jt. Ex. 50, at 23. Consequently, they recommended to the Civil Service Commission that the physical capability test be made part of the ranking of job candidates. This recommendation was adopted by the Commission in May 1980. There is no indication in the record before the Court that possible greater adverse impact upon women from a scored physical exam was considered.

At approximately the same time, a new firefighter physical examination was proposed for administration in 1980. Development of the test and the Ingram job analysis occurred

simultaneously; the job analysis did not precede test development. Tr. 57-59. With one change--the elimination of a furniture push event--the 1984 examination was identical to that administered in 1980. Three of the six events on the 1984 physical test: the beam walk, stairway climb, and sandbag carry, had previously been used in 1975 through 1978 on a pass/fail basis. The stairway climb and sandbag carry were now to be timed and scored. One event, the manual dexterity test, had been suggested by the Fire Training Academy. Jt. Ex. 50, at 1. The remaining two events: the pike pole pull and the equipment hoist were new in 1980. They appear to have been developed as simulations of firefighting tasks; both are tests primarily of upper body strength.

In summary, the 1984 firefighter physical examination evolved from previous physical examinations with the addition of several events

to test upper body strength. In three respects, test developmen departed from reasonable professional standards and practices. First, despite the emphasis upon endurance and agility in the Battelle study, events were not developed to test specifically for these abilities. Second, it is both striking and surprising that the various physical tests since 1975 were never administered to incumbant firefighters in any systematic way. This means that, to a large extent, test development proceeded in a vacuum. Third, a major change in approach from a pass/fail to a scored physical examination occurred in 1980 without any apparent consideration being given to possible greater adverse impact upon women. This change was purportedly justified by the Ingram job analysis, which will be examined in detail in the next section of this Opinion. In broader perspective, the City had readopted the approach used in 1973 prior to

Dozier and Brandt after having experimented briefly with alternative approaches in 1975 and 1978. It appears that Dozier and Brandt had caused little change of approach.

The record contains relatively little detail regarding the development of the mechanical reasoning test used in 1984. Dr. Gerald Barrett, an industrial psychologist, testified that he developed a written test for entry-level firefighters for the City of Akron in 1974 connection with employment discrimination litigation. The test has been used since then under court supervision. Tr. 652-653. The City of Columbus adopted Barrett's test for use in 1980. Jt. Ex. 16, at 1-7.

Job Analysis

The Uniform Guidelines require that any validity study, i.e., any demonstration of the job-relatedness

of a selection procedure, should be based upon a jo analysis, that is, "a review of information about the jo for which the selection procedure is to be used." 2 C.F.R. §. 1607.14(A). The job analysis need not b conducted by any particular method, provided that i yields the information required for the specifi validation strategy used. Id. In the instant case defendants have, of necessity, relied primarily upo content validity studies. Regarding such studies, th Guidelines state:

> There should be a job analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative importance ... Any job analysis should focus on the work behavior(s) and the tasks associated with them.... The work behavior(s) selected for measurement should be critical work behavior(s) and/or important work behavior(s) constituting most of the job.

29 C.F.R. §. 1607.14(C) (2).

In addition to the Battelle study discussed supra, the City has conducted two analyses of the job of firefighter - the 1980 Ingram and Kriska job analysis ("Ingram report"), Jt. Ex. 18; and a report prepared specially for purposes of this litigation by Landy, Jacobs and Associates ("Landy study"), Jt. Ex. 17. There is no issue in this case regarding the adequacy of the job analyses. Plaintiffs' expert Dr. Cranny testified that he saw no major problems with these two analyses. Tr. 289, 314-315. Based upon this testimony and its own examination of the relevant exhibits, the Court concludes that the City has complied with the requirements of the Guidelines. Accordingly, these job analyses will be summarized in this section of the Opinion only to the extent necessary to evaluate the validity studies performed by the City, which is the subject of the next section of this Opinion.

The research underlying the Ingram report was conducted in five stages. In the first stage, a comprehensive list of job tasks was compiled through questionnaires given to incumbent firefighters. At stage two, this list of job tasks was shown to a sample of firefighters who were asked to score the tasks according to frequency of occurrence, consequence of error and probability of error. Based upon these scorings by firefighters, a ranking of job tasks by "task value" was derived. The highest "value" tasks, which were ranked "5", were those that occurred frequently, where error was likely, and, if an error occurred, it would likely have serious consequences. The lowest "value" tasks were scored one.

The Court has examined this aspect of the Ingram report in detail. It appears that firefighters rank as most important-and consequently scored "5"--job tasks involving judgment and safety

procedures. Of the fourteen tasks ranked "4", only several: manipulating and working from ladders, and immediate response, appear to be predominantly physical in character. None of these physical tasks was directly simulated in the scored part of the physical test in 1984. Most of the tasks ranked "4" appear to involve primarily judgment and safety.[4]

In stage three of the study, Ingram presented firefighters with the task list she had formulated, with each task tentatively matched with knowledges, skills and abilities ("traits") needed to perform the task. The firefighters were asked to add or delete traits necessary to perform the job. Eleven traits were added. The product of this stage of the research was a matching of tasks and traits. In his testimony at trial, Dr. Cranny questioned this attempt to determine the abilities required by various tasks. Tr. 290-291. Having examined the Ingram report, the Court is

satisfied with the trustworthiness of these inferences. The abilities are described concretely, and in terms that appear comprehensible to ordinary persons. The Court is satisfied that incumbent firefighters can make reasonable judgments about the knowledges, skills and abilities they must use daily.

In stage four of the research, firefighters were asked to rank these "traits" on three bases: the extent to which the trait must be possessed by a firefighter to perform on a "barely acceptable" level; the extent to which the trait may distinguish a superior from an average firefighter; and the extent to which applicants may be expected to possess the trait. In order to gain some idea of the beliefs of incumbent firefighters about what traits are indications of superior firefighting ability, the Court compiled a list of all traits that received a score of 2.2 or higher on the

three point scale employed in the study. The cut-off of 2.2 was chosen arbitrarily to include a reasonable number of traits; 17 of 142 traits were scored 2.2 or above.[5] Most of these traits are knowledges of various sorts. Also, traits include abilities to function under adverse environmental conditions or to deal with stress. None of these traits appears to involve, in any direct fashion, physical abilities. The sole apparent exception is remaining oriented and functioning without sight.

However, when attention is turned to those traits that firefighters regard as necessary to a barely adequate job performance, a different picture emerges. These results are reported on a scale of "0" to "1", with "1" representing "yes" and "0" representing "no." Nineteen traits received scores of "1.000."[6] These abilities are predominantly, although not exclusively, physical in nature. Thus, the Ingram report tends to support the conclusion

that, while physical abilities are highly important as minimal qualifications, they are not particularly good indicators of superior firefighting ability.

In the fifth and final stage of the research, ratings were developed to indicate the relative importance of each trait. These ratings were derived by summing the values--as determined at stage two of this project--of the tasks that were associated with the particular trait. To gain an understanding of the judgments of firefighters, the Court examined all traits that were awarded task value scores greater than or equal to twenty.[7] Again, knowledges of various kinds, e.g., of proper lifting techniques, tend to predominate on this list. The physical ability to use fire department equipment, such as pike poles, also ranks prominently on this list. On the whole, however, this section of the study reinforces the general conclusion that knowledge

distinguishes the better firefighter.

Ingram also sought to determine what traits should be considered for testing purposes. A trait was inappropriate for testing purposes if it could only be learned on the job or if all applicants already possessed the trait. Of particular interest are those traits that were labelled "degree" traits. These were those traits that not only met the minimum criteria for inclusion in a test, but also tended to indicate superior workers. Ingram recommended that any test should measure for the amount of the degree trait that each applicant possessed. Jt. Ex. 18, at 7-8. In all, there were thirty-three such degree traits. Physical traits were prominent, comprising eighteen of the total. However, with several exceptions, the task values associated with these physical degree traits tended to be relatively low.[8] The exceptions are: physical ability to use firefighting equipment, such as pike poles

and axes; ability to crawl on hands and knees; and physical ability to work from ladders.

The report draws a number of conclusions regarding testing for physical abilities.

> In many instances, a superior worker would possess more of a given physical ability than other workers. That is, the ability to lift more, work longer or climb faster was generally the mark of a better fighter.

Id., at p. 10. The report also recommended ranking applicants with respect to physical abilities.

> [S]ince in many instances the possession of a higher degree of a physical ability is better, there is justification for awarding points for the performance of certain physical activities and including the physical capability scores in the ranking process.

Id. The report recommends that candidates be ranked in the 1980 examination process on the basis of their combined scores on the written and physical test.

It appears that these conclusions are, at least in part, unsupported by the report. It is an overstatement to assert that superior physical ability was generally a mark of a superior firefighter. Rather, as noted above, firefighters themselves rated knowledge and judgment much more highly in evaluating superiority of a firefighter. Further, the conclusion that ranking by physical test scores is appropriate neglects the value of the tasks associated with various physical abilities. The associated task values vary widely; in many cases, they tend to be quite low. Tr. 295–303; Plaintiff's Ex. 2.

Thus, the Court concludes that the Ingram report only weakly supports one of its central conclusions: that superior physical ability distinguishes superior from average firefighters. Despite this failing, the Ingram report contains substantial, detailed

information about the job of firefighter in Columbus, and a wealth of information about firefighters' understanding of their job.

The second job analysis, the Landy report, was designed specifically to be used as basis for a validation study, which will be discussed in the following section. It contains less detailed information and a higher degree of aggregation of data because of this linkage to a specific validation strategy. In this section of the Opinion, it will be discussed briefly as background to the Landy validation study, and to resolve certain factual issues that have arisen.

Landy began his job analysis with a list of tasks that was aggregated into twenty-eight functional categories. Examples are: firefighting - operates and advances hose lines and fire extinguishers; forcible entry - pries open or breaks down doors or windows using appropriate tools while wearing full firefighting gear; and extrication - extricates victims

from buildings or cars using appropriate tools. Jt. Ex. 17, App. B. This list of grouped tasks was then shown to incumbent Columbus firefighters, who were asked to rate its verisimilitude on a one (for poor) to five (for very good) scale. The incumbents awarded an average score of 4.06. Id., p. 6. Landy concluded that the task list was a good representation of the job. Id., p. 6. No attempt was made, however, to correct the list in light of the responses of incumbent firefighters.

Next, Landy asked incumbent firefighters to rate the importance of the various task groups. They were asked to distribute one hundred points among the various groups to reflect their importance in preserving life and property. Jt. Ex. 17, App. E. This yielded an average importance score for each of the twenty-eight functional tasks groups. The most highly

rated tasks, and their associated scores were: firefighting (8.1), rescue (7.3), search (5.9), emergency medical treatment (5.8), driving (5.0), engine operation (4.4), apparatus operation (4.3), and extrication (4.1). Firefighters were also asked to rate the various task groups according to how frequently they were performed. Id., App. G, H. The most important tasks, in terms of saving life and property, tend to occur infrequently, whereas less important tasks, e.g., equipment maintenance, occur daily. Landy elected, in light of this inverse relation between importance and frequency, to ignore frequency for the remainder of the study. This judgment is questionable, for the two measures could have been combined. Ingram had done so in her job analysis. However, this does not appear to the Court to be a fundamental problem with the Landy job analysis.

The next step in the Landy job analysis is more controversial. Landy sought to determine the physical abilities that were necessary to perform the various job tasks. To do this, he used a taxonomy of human abilities developed by Dr. Edwin Fleishman, an industrial psychologist. Fleishman sought to devise a list of abilities that underlay all human performance; the list was to be comprehensive and its elements were to be independent of one another. The list of abilities was created by reliance on a statistical technique, factor analysis. As this suggests, the Fleishman abilities are abstract concepts that are linked to a theory of human performance.

For instance, Fleishman distinguished three kinds of strength. Static strength refers to the amount of force that a person can exert against an immovable or very heavy object. This is similar to the everyday concept of strength. Explosive strength, on the other

hand, refers to the ability to use energy in one or a series of explosive muscular acts. An example would be the strength used in jumping over a barrier. Dynamic strength is the ability to use one's arms and trunk repeatedly to move one's body weight over a distance, e.g., climbing a rope. The Fleishman terms are not everyday ones and it was necessary to training firefighters in their meaning. Further, the distinctions drawn in the Fleishman classification are not common-sense ones, as is illustrated by the three kinds of strength.

Landy asked incumbent firefighters to rate the extent to which the various Fleishman abilities were involved in performance of the tasks involved in the various task groups. For this purpose, the twenty-eight task grouping previously defined were aggregated into sixteen groups. Id., App. J. Prior to making the ratings, the firefighters were instructed in the

Fleishman classification, and discussions were held.[9] Then, the firefighters completed the task ratings by distributing one hundred points across the various Fleishman abilities to reflect their relative role in performance of a particular group of tasks. The scores assigned to each ability were then averaged across the various task areas. Finally, values were recalculated to reflect the relative importance of each of the task areas, as previously determined.

At trial, plaintiffs' expert Dr. Cranny criticized the reliance upon the Fleishman abilities in the Landy study. He testified that the inferences about the abilities involved in a task were inherently unreliable, even when made by job incumbents. Tr. 321-323; Jt. Ex. 10. Concerns about whether firefighters understood the abstract categories of Fleishman's taxonomy were also expressed by the court in Berkman v. City of New York, 536 F. Supp. 177, 189-190

(E.D. N.Y. 1982), aff'd, 705 F.2d 584 (2nd Cir. 1983), where the Fleishman taxonomy had also been used in the job analysis. In the instant case, although the Court feels some skepticism about reliance on terminology so distant from ordinary experience, the use of the experience, the use of the Fleishman abilities does not appear to cause major problems. Landy calculated intra-class correlations for firefighters using the Fleishman categories. This statistic measures the amount of agreement of the various individuals about the extent to which the various task groups involve a particular ability. In the case of firefighters, the intra-class correlation was quite high, equalling .95. Jt. Ex. 13; Tr.982-983. In addition, Dr. Landy also presented at trial certain exhibits summarizing the judgments of firefighters about particular tasks and abilities. These exhibits showed a good deal of variability and discrimination,

suggesting that the firefighters were making reasonably accurate judgments. Tr. 984-987.

The results of the abilities analysis of incumbent firefighters are presented in Appendix M of Jt. Ex. 17. One conclusion reached in the report is that physical abilities account for one half of the job of firefighter. That is, of ratings assigned to all thirty-five Fleishman physical and cognitive abilities, the ratings assigned to physical abilities amount to about 50 of a possible 100 points. The seven physical abilities rated highest by firefighters with their accompanying ratings are: stamina (8.17); static strength (8.11); explosive strength (4.86); dynamic strength (4.81); multi-limb coordination (2.74); manual dexterity (2.67); and gross-body coordination (2.63). Together, these abilities account for 63% of the total physical ability composition of the job of firefighter. This information constitutes the basis for Landy's attempt to validate the 1984 physical examination.

As noted in the introduction to this section of the Opinion, there was no real dispute regarding the job analyses that had been performed by the City. It is plain that adequate descriptions of the job of firefighter in Columbus have been formulated. Several general conclusions can be reached here. First, neither the Ingram report nor the Landy report justifies the conclusion that possession of more of a particular ability is, in all circumstances, better. For all these reports show, it may well be true that a firefighter requires enough of a particular ability to do the job well, and that any more of that ability is merely redundant. Further, neither report contains any data on how quickly firefighters perform particular tasks. There is much conflicting testimony in the record about the speed at which firefighters work; neither of these reports addresses this issue.

Test Validation Studies

In this section, the Court summarizes the two test validation studies undertaken by the City. The first was authored by Dr. Kriska and Constance Hines and was intended to fulfill one of the requirements of this Court's Order in Dozier v. Chupka, supra. Jt. Ex. 16 ("Kriska/ Hines report"). Although it is based upon data from the 1980 firefighter examination, it is relevant to the 1984 examination by virtue of the substantial overlap between the two examinations. The second study was that undertaken by Landy, Jacobs and Associates specially for purposes of this litigation. It involved solely an analysis of the 1984 physical test. Jt. Ex. 17 ("Landy report").

Validation refers to the process of gathering evidence to show the job-relatedness of a test or selection device. Validity may be demonstrated by different kinds of studies: criterion related studies, content validity studies, or construct validity studies.

29 C.F.R. §. 1607.5(A); Harless v. Duck, 619 F.2d 611, 616 n.5 (6th Cir.), cert. denied, 449 U.S. 872 (1980). Only the first two approaches are relevant to this litigation. The Kriska/Hines report is primarily a criterion-related validity study. The Landy report, on the other hand, is an example of a content validity study.

In the criterion-related validity study, an attempt is made to collect data to show that the test predicts important aspects of actual job performance. In such a study, thus, evidence is sought to show the association of test scores and measures of actual performance n the job, the criteria. There are two kinds of such studies. In a predictive validity study, an applicant's test scores and subsequent performance on the job as an employee are compared. In a concurrent validity study, on the other hand, the test scores of present employees are compared with their present job

performance. Both approaches were used in the Kriska/Hines study. Where it is possible, a criterion-related study is preferable, because it is the most direct approach to showing job-relatedness. However, due to problems with measuring job performance, a criterion approach is not always feasible. Tr. 241-243.

In a content validation study, evidence is gathered to show that the content of the test, i.e., the questions or tasks comprising the test, are representative of the content of the job, i.e., the important or critical tasks comprising the job. An attempt is made to determine the degree to which test items are representative of the job. 29 C.F.R. §. 1607.14(c)(4); Jt. Ex. 44, at 11. Although a content validation approach is less direct than a criterion-related approach, it is nonetheless a permissible method for demonstrating validity. Firefighters Institute for Racial Equality v. City of St.

Louis, 549 F.2d 506, 511 (8th Cir. 1977). It should also be pointed out that criterion-related approaches and content approaches are not mutually exclusive in any respect; they are simply different strategies for collecting evidence regarding job-relatedness. Jt. Ex. 44, at 9-11; Tr. 243.

As noted above, the Kriska/Hines report sought to demonstrate criterion-related validity by both predictive and concurrent studies. Kriska/Hines used two categories of variables as measures of on-the-job performance. Two measures were derived from ratings of firefighters by their supervisors based upon observation of the firefighters over a period of time. These two scales were measures of performance at the fire scene, and overall performance. Jt. Ex. 16, at 4-2 to 4-14. The other category of measures of job performance derived from testing programs that are used in the Fire Division to evaluate training success. One was the Training Academy Final Average, a

composite score consisting of instructor ratings and scores on written exams during initial firefighter training. The other training measures were written examinations used in post-Academy training; these are called the Firefighter I, Firefighter II and Journeyman examinations. Id., at 4-14 to 4-18.

In the predictive study, Kriska/Hines sought to find significant and substantial correlations between scores upon the physical capability test and these criterion measures. The results were disappointing. The correlation of the physical test scores with supervisor's ratings of performance at the fire scene was .00, that is, there was no association at all. The correlation of the physical test with supervisor's overall ratings was -.03, that is, there was a very slight negative association. Jt. Ex. 16, at 5-19. The only statistically significant correlation found with the training measures was with the Training Academy

Final Average; this correlation was .32. Neither the Firefighter I or II examinations were significantly correlated with the physical test. Id.

Kriska/Hines noted a number of statistical problems that might be causing their reported correlations to underestimate the actual correlation.[10] Jt. Ex. 16, at 5-27. Dr. Landy corrected statistically for these problems, and recalculated the correlations between physical test scores and the various criterion measures. Landy's calculations have the effect of doubling the correlation between the physical test and the Training Academy Final Average; it is variously reported as being between .60 and .72. Defendant's Ex. E, Tables 1, 4, 6. Otherwise, nothing else changes, that is, all other attempted correlations with criterion measures remain nonsignificant. The correlations with supervisor's ratings remain essentially zero, as before. Id., Tables 1, 3, 5.

Kriska/Hines also report results of a predictive study of Barrett's mechanical test, which is also under challenge in this litigation. They found significant correlations of mechanical test scores with Training Academy Final Average (.54) and the Firefighter I examination (.38). Jt. Ex. 16, at 5-19. Correlations with the performance measures were not significant. As before, Landy's recalculations tended to increase these reported correlations somewhat.

Kriska/Hines did not examine the physical test in their concurrent validity study. They expressed the view that it was possible that training and performance of firefighting tasks made applicants and incumbent firefighters different from one another. Also, administering the test to incumbents would increase the cost of the study. Jt. Ex. 16, at 6-4. Thus, the City was no more willing to administer its physical test to incumbent firefighters to validate its test than it was in the process of developing the test.

The Kriska/Hines study did report the results of administering the Barrett mechanical test to incumbent firefighters. The scores of incumbents were then correlated with certain of the criteria measures previously discussed. There were significant correlations as follows: with Training Academy Final Average: .53; with Firefighter I examination: .31; with supervisor's ratings of performance at the fire scene; .30; and with supervisors' ratings of overall performance: .32. Jt. Ex. 16, at 6-17. This section of the report concluded that the mechanical aptitude tests were significantly correlated with training success and on-company performance and, therefore, should be retained as part of the test for selecting firefighters. Id., at 6-29.

The other validation study was the Landy study, which pertained solely to the physical examination. This study was an attempt to demonstrate job-relatedness through a content validation strategy.

In the previous section, Landy's job analysis was discussed. As will be recalled, the culmination of that analysis was a rating, by incumbent firefighters, of the relative importance of the Fleishman physical abilities in the job as a whole. This rating had been derived by averaging across the various task groups formulated in the job analysis, and weighting for importance of the tasks. Landy's validation strategy was direct, yet elegant. He asked a group of industrial and organizational psychologists to make an evaluation of the 1984 firefighter test similar to that made by the firefighters of the job. Like the firefighters,the psychologists were given the Fleishman abilities with explanatory and illustrative material, and were presented with information, including a videotape, about the 1984 firefighter test. For each event, they

were asked to distribute one hundred points across the various abilities to reflect the extent to which the ability was tested by the particular event. These results were then averaged across the various events to yield an overall measure of the extent to which a given ability was important in the 1984 examination. Values were recalculated to omit the beamwalk event; this is reasonable because only one person failed that event.

The seven highest rated abilities accounted for approximately 80% of the total points awarded by the psychologists to all eighteen abilities. These highest rate abilities were:

Ability	Score*	Top 7 in *** Firefighters Rating	Score in ** Firefighters Rating
Speed of Limb Movement	17.9	no	2.4
Dynamic Flexibility	14.5	no	3.8
Static Strength	13.4	yes	16.2
Explosive Strength	10.0	yes	9.8
Stamina	9.8	yes	16.4
Manual Dexterity	7.7	yes	5.4
Wrist-Finger Speed	6.1	no	1.8

*Source: Jt. Ex. 17, App. R.

**Source: Jt. Ex. 17, App. M. Because physical abilities were assigned only 49.9 points out of a possible total of 100 points, the scores reported in Appendix M are doubled for purposes of comparison with the results reported in Appendix R. The remaining points in Appendix M were assigned to cognitive abilities.

***The seven highest rated physical abilities in Appendix M with their associated (corrected) scores were: Stamina (16.4), Static Strength (16.2),

Explosive Strength (9.8), Dynamic Strength (9.6), Multi-Limb Coordination (5.4), Manual Dexterity (5.4), and Gross-Body Coordination (5.2). These seven abilities account for 63% of the total physical points awarded by firefighters.

From the Landy study, it is possible to offer a qualitative appraisal of the job-relatedness of the 1984 firefighter test from the point of view of the underlying physical abilities purportedly measured. First, the test appears to overemphasize certain abilities. This is quite extreme in the case of speed of limb movement and dynamic flexibility. (Dynamic flexibility is defined as the ability to make repeated arm or leg flexing movements with some speed, e.g., pulling in a rope, hand over hand. Jt. Ex. 17, App. I.) There is also substantial overweighting of wrist-finger speed in the test. This overweighting appears to result from the timed nature of the test. The emphasis on extreme speed that, of necessity, character

izes such a timed test does not appear to be reflected in firefighters' appraisals of their jobs.

The test also underweights certain abilities that were thought to be important by firefighters. This is most striking in the case of stamina, the physical ability most highly rated by firefighters. Dynamic strength also appears to be under weighted, having been rated 5.8 by the psychologists, but a higher (corrected) 9.6 by the firefighters. Finally, the 1984 test appears to have achieved a reasonable fit with the static strength, explosive strength and manual dexterity required by the job; the relative ratings appear sufficiently comparable to justify this conclusion.

Dr. Landy testified that there was a "good match" between the test and the job. Tr. 961. He based this conclusion upon the observation that

the abilities most highly rated by the firefighters: endurance, static strength, explosive strength, and dynamic strength, were also important in the test. Tr. 1010; Jt. Ex. 17, at 21. Dr. Cranny disagreed with this conclusion. He calculated a correlation coefficient to measure the extent to which the relative abilities for the job and that the test were rated in the same order by the respective judges. The correlation was .45, a "rather low degree of correspondence." Tr. 338; Plaintiffs' Ex. 1. Although this calculation was questioned by Dr. Landy, Tr. 1009, it appears reasonable. Tr. 336-339. It appears that the experts are choosing to characterize the fit between test and job in different ways, rather than contradicting one another. The test does reflect certain abilities that are important to the job; this is especially true in the case of the various kinds of strength. On the

other hand, there are other abilities that are not reflected in the test proportionally to their apparent importance in the job, and yet other abilities that are overemphasized in the test. The controlling question is whether the degree of fit achieved by the test is sufficient; this question will be addressed in the following section.

Legal Discussion

The Court having concluded that plaintiffs have demonstrated adverse impact from he 1984 firefighter examination, it becomes the defendants' burden to show that the test "bears a manifest relationship to successful and efficient job performance." Harless v. Duck, 619 F.2d 611, 616 (6th Cir.), cert. denied, 449 U.S. 872 (1980); Griggs v. Duke Power Co., 401 U.S. 424, 432 (1971). The test of manifest relationship looks to whether the discriminatory employment

practice is "necessary to safe and efficient job performance." Chrisner v. Complete Auto Transit, Inc., 645 F.2d 1251, 1252 (6th Cir. 1981). "Necessary" here does not mean indispensable, but rather "substantially promote[s] the proficient operation of the business." Id. Nonetheless, manifest relationship is intended to set a "high standard." E.E.O.C. v. Ball Corp., 661 F.2d 531, 541 (6th Cir. 1981). If, but only if, the employer meets the burden of establishing manifest relationship, the burden shifts back to the plaintiff to show that there is an alternative selection device with less disparate impact that would also serve the employer's legitimate interests. Chrisner v. Complete Auto Transit, Inc., supra, at 1263. Because the Court concludes that the defendants have not met their burden, the issue of alternative selection devices is not reached here.

In the instant case, the Court concludes that the defendants must demonstrate manifest relationship by showing that the 1984 test was validated in conformity with the standards set forth in the Uniform Guidelines, 29 C.F.R. §s. 1607.1 et seq. Although the Guidelines are not substantive regulations promulgated as law, they are entitled to "great deference." Albermarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975). The Guidelines have been followed by those courts that have decded cases involving discrimination in testing in fire departments. Berkman v. City of New York, 536 F. Supp. 177 (E.D. N.Y. 1982), aff'd, 705 F.2d 584 (2d Cir. 1983); Fire Institute for Racial Equality v. City of St. Louis, 549 F.2d 506, 510-511 (8th Cir.), cert. denied, 434 U.S. 819 (1977); Vulcan Society v. Civil Service Commission, 360 F. Supp. 1265, 1273 n. 23 (S.D. N.Y.), mod., 490 F.2d 387 (2d Cir. 1973). An additional reason for applying the Guidelines is

that in Dozier v. Chupka, 395 F. Supp. 836 (S.D. Oh. 1975), this Court ordered the City to validate its hiring criteria for firefighters in compliance with the Guidelines. Id., at 859-860. In so concluding, the Court is mindful that the Guidelines are meant to be consistent with professional standards for testing, and that these are not unchanging. 29 C.F.R. §. 1607.5(C), (A). Thus, the Court considers it appropriate to consider also the standards set forth in Standards for Educational and Psychological Testing, published by the American Psychological Association in 1985 ("Division 14 Standards"). Jt. Ex. 44.

Relying upon Spurlock v. United Airlines, Inc., 475 F.2d 216 (10th Cir. 1972), the City argues that it should be held to a lower quantum of proof of job-relatedness because the job of firefighter implicates public safety. In Spurlock, the Court held:

> when the job clearly requires a high degree of skill and the economic and human risks involved in hiring an unqualified applicant are great, the employer bears a correspondingly lighter burden to show that his employment criteria are job-related.

Id., at 219. The Sixth Circuit adopted this doctrine in Chrisner v. Complete Auto Transit, Inc., supra. Subsequently, the Court of Appeals explained that the doctrine was restricted to the "narrow catergory of jobs which greatly implicate human safety, e.g., airline piloting and over-the-road trucking." E.E.O.C. v. Ball Corp., supra, at 541 n. 20.

For the following reasons, the Court concludes that the Spurlock doctrine does not alter the defendants' burden of showing compliance with the Guidelines. First, the Spurlock Court relied upon an E.E.O.C. regulation, then existing 29 C.F.R. §. 1607.5(c)(2)(iii) as the basis for its holding.

However, when the Guidelines were revised in 1978, this provision was not included. The natural assumption is that this provision has been incorporated into or superseded by the standards presently set forth in the Guidelines. In addition, the Spurlock doctrine has been applied mainly in cases involving education or experience requirements or other non-scored objective criteria. See B. Schlei and P. Grossman, Employment Discrimination Law (2d ed. 1976), at 167-173. It has been frequently been applied in cases involving scored tests, where distinct standards have been developed by the courts.

Turning now to the merits of defendants' case, the Court first concludes that defendants have met their burden with respect to the mechanical reasoning test. There is no evidence before the Court about the content of this test as administered in 1984. The plaintiffs have done little more than raise the

issue by showing adverse impact; they virtually abandoned the claim at trial and in their brief. Nevertheless, the defendants have produced evidence of validation of this test, which, under the circumstances, the Court can only conclude is sufficient to meet their burden.

Defendants' evidence is of two sorts. First, in the Kriska/Hines concurrent validation study, significant correlations with both training performance and supervisor's ratings of on-the-job performance were shown. Jt. Ex. 16, at 6-17. Plaintiffs' contention that the correlations in this study are too low to validate the test is unpersuasive. See, e.g., B. Schlei & P. Grossman, supra, at 129; also 1983-1984 Supp., at 18. The Guidelines set no minimum standards for correlation coefficients in criterion-related studies. 29 C.F.R. §. 1607.14(B)(6). The Court considers it appropriate to rely upon both training performance and

on-the-job performance as validating criteria. Mechanical reasoning ability, it would appear, is necessary both to successful completion of training and performance on the job.

In addition, defendants presented testimony from Dr. Gerald Barett, the developer of the mechanical reasoning test. He testified that the test had been developed to use in selecting firefighters in Akron, and that a test validation study had been performed on the tests as used there. Tr. 652-657; Jt. Ex. 19. He further testified that, based on his review of data about the job of firefighter in Columbus, his knowledge of the job of firefighter in Akron, and the general literature on firefighting, the job of firefighter was similar in both cities. Tr. 658, 660. This testimony is uncontradicted. In fact, plaintiffs' expert,

Dr. Magel, testified that evidence regarding firefighting in one city was applicable to another city, in his words, "firefighting is firefighting." Tr. 461. The Court concludes that requirements of the Guidelines for reliance upon validity studies conducted by other users have been met. 29 C.F.R. §. 1607.7.

The Court further concludes that defendants have failed to show that the 1984 firefighter physical examination is valid by means of the Kriska/Hines predictive criterion-related validity study. That study found a significant and substantial correlation between physical test scores and the Training Academy Final Average, but no other meaningfully correlation with the other training measures or, more important, with on-the-job measures. In Dozier v. Chupka, supra, this Court rejected a contention that a correlation of test scores and training academy scores was sufficient to

validate a test. Id., at 853. The Court sees no reason to abandon this proposition here. It is true that the Supreme Court has held that a positive correlation of a test with training course performance may be enough to validate a test apart from a possible relationship to on-the-job performance. Washington v. Davis, 426 U.S. 229, 250 (1976). However, courts of appeals have interpreted this holding to apply only in the case of minimal standards necessary to successful completion of a training program. Guardian's Association v. Civil Service Commission, 633 F.2d 232 (2d Cir. 1980), aff'd, 463 U.S. 582 (1983); Ensley Branch of NAACP v. Seibels, 616 F.2d 812, 819-822 (5th Cir. 1980); Craig v. County of Los Angeles, 626 F.2d 659, 662-663 (9th Cir. 1980). Physical ability, of course, is not something that is merely needed to train as a firefighter; it is necessary on the job. Thus, the absence of any non-zero correlations with on-the-job measures of

performance is fatal to any claim of criterion-related validity.

The more important issue, to which the parties have devoted the most attention, is whether the defendants have shown that the 1984 physical examination is content-valid. More particularly, the controlling question is whether the Landy study, Jt. Ex. 17, constitutes such a demonstration. Plaintiffs raise a number of objections to the design and execution of the Landy study: the study improperly relied upon abstract physical abilities; the study failed to provide operational definitions; and the ratings of relative importance of various physical abilities by firefighters and psychological were unreliable. Plaintiffs also contend that, even ignoring these alleged problems of the study, it does not demonstrate content validity for two reasons: the test events did not accurately reflect the complexity of actual job tasks, and the

results of the Landy study did not show sufficient proportionality between test and job to permit rank-ordering of applicants on the basis of test scores. The Court concludes that, except for the last, these contentions are without merit. However, the last point, by itself, compels the conclusions that the 1984 physical exam is not content-valid and, therefore, its use in the 1984 firefighter selection process constituted impermissible discrimination.

Plaintiff's objection to the design and execution of Landy study itself are readily disposed of. The Guidelines expressly permit selection procedures that measure knowledges, skills or abilities to be justified by content validity. 29 C.F.R. §. 1607.14(C)(1). However, they require that the knowledge, skill or ability be operationally defined. 29 C.F.R. §1607.14(C)(4). The Court concludes that this

requirement was met by the Landy study. Jt. Ex. 17, App. I. Finally, for reasons discussed supra, the Court concludes that judgments of firefighters and psychologists are not so unreliable as to undermine the study.

Plaintiffs also object that the test events fail to approximate actual job tasks. The Guidelines provide:

> [T]o be content valid, a selection procedure measuring a skill or ability should closely approximate an observable work behavior ... If a test purports to sample a work behavior ..., the manner and setting of the selection procedure and its and complexity should closely approximate the work situation.

29 C.F.R. §. 1607.14(C)(4). On the whole, the 1984 firefighter examination was a reasonable approximation of the actual tasks.

The more telling objection to the 1984 physical examination is not that the events comprising

it fail to approximate actual job tasks; it is that, taken as a whole, the test fails to reflect accurately the content of the job. The Guidelines provide:

> A selection procedure can be supported by a content validity strategy to the extent that it is a representative sample of the content of the job.

29 C.F.R. §. 1607.14(C)(1). The Division 14 Standards also speak of representativeness. Jt. Ex. 44, at 10–11. This has been interpreted to require that a test, to be content valid, must reflect all or nearly all the important aspects of the job. Firefighters Institute for Racial Equality v. City of St. Louis, 549 F.2d 506, 511–512 (8th Cir.), cert. denied, 434 U.S. 819 (1977); accord, Guardian's Association v. Civil Service Commission, 630 F.2d 79, 98–100 (2nd Cir. 1980), cert. denied, 452 U.S. 940 (1981); Berkman v. City of New York, supra, at 195; Burney v. City of Pawtucket,

559 F. Supp. 1089, 1101-1103 (D. R.I. 1983); see generally, B. Schlei and P. Grossman, supra, at 130 n. 135-137. Based upon the Landy study, the Court concludes that the 1984 physical examination does not meet this standard of representativeness. As discussed in the preceding section, the test overemphasizes speed of limb movement and dynamic flexibility, while it underemphasizes endurance. The fact that the test appears to reflect, more or less accurately, the strength necessary for the job does not, by itself, validate the test. Although there was testimony at trial that stronger firefighters would have more endurance, Tr. 970-972, such generalized testimony cannot be accepted as a substitute for concrete evidence based upon a job analysis.

In addition, if a test is to be used to rank-order applicants, it must be more than merely content valid. The Guidelines provide:

> If a user can show, by a job analysis or otherwise, that a higher score on a content valid selection procedure is likely to result in better job performance, the results may be used to rank persons who score above minimum levels. Where a selection procedure supported solely or primarily by content validity is used to rank job candidates, the selection procedure should measure those aspects of performance which differentiate among levels of job performance.

29 C.F.R. §. 1607.14(C)(9). The Guidelines recognize that a test which may be valid as a pass/fail test, may not be valid as a ranking test, because of likely greater adverse impact. 29 C.F.R. §. 1607.5(G).

The courts have followed these special requirements for ranking tests. In Williams v. Vukovich, 720 F.2d 909 (6th Cir. 1983), the Court of Appeals stated:

> Ranking is a valid, job-related selection technique only where the test scores vary directly with job performance.

Id., at 924, citing Guardian's Association of New York v. Civil Service Commission, 630 F.2d 79, 100 (2nd Cir. 1980), cert. denied, 452 U.S. 940 (1981). In Guardian's Association, the Second Circuit had held:

> Permissible use of rank-ordering requires a demonstration of such substantial test validity that it is reasonable to expect one- or two-point differences in scores to reflect differences in job performance.

Id., at 100-101. So far as the Court's research discloses, this appears to be the unanimous view of the courts. See generally, B. Schlei & P. Grossman, supra, at 155 n. 17; 1983-1984 Supp. at 18 n. 42.

In Berkman v. City of New York, 536 F. Supp. 177 (E.D. N.Y. 1982), aff'd, 705 F.2d 584 (2nd Cir. 1983), the Court struck down a physical test for firefighters on a number of grounds, among them that

it was insufficiently precise to justify ranking of candidates. Id., at 210-212. The Court objected especially to the premium placed on maximum speed and all-out effort on the test, which, like the instant test, was timed. The Court concluded that such a test failed to reflect the actual demands of firefighting which, in many circumstances, requires endurance and pacing. Id., at 212. The court is aware of no case, and the defendants cite none, in which a ranking test has been upheld as a selection device for firefighters. See also Firefighters Institute for Racial Equality v. City of St. Louis, 616 F.2d 350, 357-360 (8th Cir. 1980).

The Landy report briefly addressed the issue of the justifiability of ranking candidates by their scores on the physical test. To determine this, firefighters were asked to estimate the level of each ability necessary to do the job of firefighter at three

performance levels: minimum competence, average competence, and outstanding performance. For this purpose, the seven most important physical abilities were selected. Groups of firefighters were asked a series of questions, of which the following is representative: Does a firefighter need to be very low, below average, average, above average or well above average in this ability to perform at an outstanding level: Jt. Ex. 17, App. S. The results are predictable, given the question format. It is hardly surprising that firefighters who are asked how much--below average, average, or above average--of an abstractly characterized ability is necessary to perform at, say, a minimally competent level, will tend to answer: below average. This exercise well illustrates the reasons why the law has developed rules against leading questions; it has no apparent bearing on any of the actual issues in this case, however.

At trial, the defendants presented testimony that firefighters frequently work as quickly as possible, going all-out to attack a fire agressively. In fact, there was a great deal of testimony at trial about firefighters working at an all-out pace versus firefighters pacing themselves. The clear import of the testimony, taken as a whole, is that sometimes firefighters work all-out, and sometimes they pace themselves; it depends on the task at hand. Anecdotal evidence regarding the speed at which firefighters must work is not sufficient to justify a timed, competitive examination. There must be systematic evidence based upon a job analysis. The Battelle researchers weighed actual pieces of firefighting equipment to determine the strength necessary to perform the job. Jt. Ex. 24, at 6-8. It is hard evidence such as this that is necessary to justify an examination with adverse impact.

The defendants also presented evidence that women, on the whole, lack the upper body strength of men, and have lower levels of aerobic capacity. This is undisputed; indeed plaintiffs' expert Dr. Magel stated at trial: "[W]e know for a fact that women perform less well in most fitness measures other than tests of flexibility or balance." Tr. 457. Firefighting is physically demanding work, defendants argue, and men are better equipped to perform this work than women. From the evidence at trial, there appears to be some truth to this. However, this argument is based upon a misconception of the role of the Court in a Title VII case. It is not the province of the Court to determine whether women should be firefighters, or how many women should be firefighters. Rather, it is the Court's duty to evaluate a test in light of the standards set forth in Title VII. How many women should be

firefighters can be decided only by the administration of a validated examination.

Accordingly, the Court concludes that the defendants have failed to show that the 1984 firefighter physical test is content valid. This conclusion is based upon two reasons: that the test, taken as a whole, does not represent the physical demands of the job, and that there is no evidence that higher scores on the test vary directly with job performance to justify ranking. Consequently, the defendants engaged in discrimination on the grounds of sex when they used the 1984 examination to firefighters can be decided only by the administration of a validated examination.

Accordingly, the Court concludes that the defendants have failed to show that the 1984 firefighter physical test is content valid. This conclusion is based upon two reasons: that the test, taken as a whole, does not represent the

physical demands of the job, and that there is no evidence that higher scores on the test vary directly with job performance to justify ranking. Consequently, the defendants engaged in discrimination on the grounds of sex when they used the 1984 examination to select firefighters. Thus, the plaintiffs are entitled to judgment on their Title VII claims regarding the 1984 physical exam.

IV.

The Court having concluded that defendants have failed to show that the 1984 firefighter physical examination was job-related, plaintiffs are entitled to relief. This relief has two aspects: prospective relief to assure future compliance with Title VII, and retrospective relief to remedy the effects of past discrimination.

The Court understands that no further hiring from the 1984 eligibility lists will occur. However, the City will at some point have to administer

new examinations and generate new eligibility lists from which future training classes will be selected. Thus, the critical aspect of prospective relief is to ensure that the next examination and selection process complies with the requirements of Title VII. Accordingly, the City will be ordered, prior to administration of any future firefighter physical tests, to modify the test so as to eliminate the problems found with the 1984 test by this Court.

Those problems are two fold: lack of representativeness, and the use of rank-ordering. To eliminate the former, the City must redesign the test so that it reasonably reflects the physical abilities actually used on the job. For this purpose, the Court may rely upon the approach used in the Landy study, and the results of the Landy analysis of the job of firefighting, as reported in Appendix M of Jt. Ex. 17. To eliminate the problems stemming from rank-ordering, the City must make a choice.

If defendants wish to continue to rank-order candidates, they must be prepared to show that rank-ordering complies with the Uniform Guidelines. Specifically, they must produce evidence to show that a higher score on the examination is likely to result in better job performance. 29 C.F.R. §.1607.14(C)(9). That evidence should be sufficient to justify any additional adverse impact that rank-ordering may have over a pass/fail test; for this purpose, defendants must determine the likely adverse impact from a pass/fail examination, the passing points of which are validated according to the standards of the Uniform Guidelines. Alternatively, defendants may choose to design and administer a pass/fail examination. In this event, the problem of representativeness of the test must be resolved, and pass points must be justified consistently with the Guidelines.[11]

Turning to the matter of retrospective relief, the fact that a non-job-related physical examination was administered by the defendants in 1984 raises the inference that there are females who would have been hired but for the discriminatory examination. However, it is impossible to determine how many of the female applicants are qualified to be firefighters, or how many, if any, would have been hired in a nondiscriminatory examination. The Court will resolve this problem by requiring the defendants to administer a nondiscriminatory examination, and to provide notice of that examination, in a form approved by the Court, to all female applicants for the 1984 examination. If women succeed upon the new examination in greater numbers than upon the 1984 examination, the Court will order the defendants to set aside a sufficient number of places in future firefighter classes to rectify any past discrimination thus revealed. On the

other hand, if female applicants do not succeed in greater numbers than before, then no set aside would be appropriate.

The parties are DIRECTED to file with the Court no later than May 23, 1986 their suggestions regarding the decree by which this remedy will be implemented. The Court will then issue its decree.

WHEREUPON, having considered the evidence and the arguments of the parties, the Court renders its decision on liability as follows: Plaintiffs have failed to prove their claim of intentional discrimination under §.1983, and their claim under Title VII with respect to the 1980 firefighter selection process; however, plaintiffs have prevailed on their claim with respect to the 1984 firefighter selection process. The Clerk shall enter JUDGMENT on liability only in favor of the defendants on the §. 1983 claim and the Title VII claim

pertaining to the 1980 examination,and JUDGMENT in favor of the plaintiffs on the Title VII claim pertaining to the 1984 examination.

IT IS SO ORDERED.

_____

UNITED STATES DISTRICT JUDGE

1

The evidence in this case comprises the testimony taken in open Court, cited to the transcript as "Tr.__"; stipulations of the parties filed in open Court, cited as "Stip.# __"; exhibits presented jointly by the parties, cited as "Jt. Ex. _"; and exhibits received into evidence on behalf of only one party, cited as, e.g., "Plaintiffs' Ex. __." In addition, the defendants made available a number of pieces of firefighting equipment for examination by the Court.

2

Defendents also assert that females passed the 1980 selection process at a rate of 89% (25 of 28), while males passed at the rate of 90% (772 of 796). The sole component of the overall testing process that was graded pass/fail was the reading comprehension test, which is not challenged in the litigation. All other components of the test were used to rank order candidates. Thus passing ratios are irrelevant.

3

Jt. Ex. 5 shows a female total score of 87.1 for Lawrence Livingston. This appears to be an error in coding for sex of applicants. The same error is repeated in Dr. Cranny's analysis, Jt. Ex. 7, at 5, and his testimony at trial. tr. 269.

4

Seven tasks were ranked "5". They were: surveying structure for possible hot spots after fire has been knocked down; using appropriate safety procedures; observing smoke and fire conditions and locating source of fire; sizing up fire and identifying appropriate extinguishing and ventilation techniques; driving apparatus according to state and local regulations; selecting shortest route to emergency scene; and maneuvering apparatus at scene to occupy best position and avoid interfering with other companies.

Fourteen tasks were rated "4". These were: locating hidden fires by seeing, feeling or smelling fire or opening walls; manipulating ladders; climbing and working from ladders with equipment or carrying people; obtaining and donning proper protective equipment; applying knowledge of heat and fluid mechanics to anticipate fire behavior; identifying and saturating potential exposures; identifying and removing flammable or hazardous materials; locating hydrant or water source with best access to fire; computing necessary line pressure; pumping water to supply hoses or sprinkler systems; responds immediately to emergency to save lives; interacts with distraught persons to obtain information; checking vital

signs of victim; and preplanning fires in industrial and commercial buildings to locate fire prevention and fighting equipment.

5

The seventeen traits and the scores they were awarded by firefighters are: knowledge of hydrant locations (2.440); knowledge of occupancy, use, and structural composition of buildings (2.379); knowledge of firefighting tactics (2.370); knowledge of methods of fire extinguishment (2.328); knowledge of safe treatment of hazardous substances (2.310); ability to receive, comprehend and follow orders (2.308); knowledge of ventilation techniques (2.283); ability to learn and improve performance (2.259); skill at remaining oriented at emergency scenes, e.g., dense smoke (2.250); ability to function without sight (2.250); knowledge of streets and addresses in district (2.241); knowledge of CFD regulations regarding positioning of apparatus at scene (2.241); knowledge of size-up procedures (2.241); ability to deal with emotional supervision (2.241); knowledge of CFD hose evolutions (2.204)); ability to put knowledge of proper use of rescue equipment into practice (2.200).

6

These traits are: physical ability to use ladders; physical ability to use hydrant wrench; physical ability to drag empty hose lines; physical ability to advance charged hose lines; physical ability to lift and operate fire extinguishers; physical ability to use CFD equipment, e.g., pike poles; physical ability to climb and work from ladders; ability to crawl on hands and

knees; ability to drag or carry adults or children; ability to detect higher temperatures by feel; ability to detect smoke or fire by smell; manual dexterity; ability to remain alert; ability to hear or read and follow instructions; physical ability to assist in loading of hose bed; ability to learn; ability to write legibly; ability to comprehend and follow orders; ability to work from heights without fear.

7

These were, in order of their scores: ability to wear mask which covers entire face (88); ability to withstand high temperatures (84); ability to function without sight (81); skill at moving around in structures weakened by fire (36); knowledge of proper lifting techniques (35); knowledge of proper use of tools (35); physical ability to use CFD equipment, e.g. pike poles (33); ability to crawl on hands and knees (29); knowledge of operation of CFD apparatus (26); communication skills-hearing and understanding speech in person (25); knowledge of search patterns used in CFD (24); knowledge of smell of materials while burning (23); communication skill--speaking (22); knowledge of proper use of CFD rescue equipment (21); ability to put knowledge of proper use of rescue equipment into practice (21); and skill at remaining oriented at emergency scenes (20). Jt. Ex. 18, Appendix 5.

8

The physical degree traits and their associated task values are: physical ability to use ladders (9);

physical ability to carry out duties of "hydrant man" (3); physical ability to use hydrant wrench (3); physical ability to drag empty hose lines (2); physical ability to advance charged hose line (4); ability to mount and operate master stream device (11); physical ability to lift and operate fire extinguishers (1); physical ability to use CFD equipment, e.g. pike poles (33); physical ability to use equipment to shore up unsound structures (3);physical ability to sue tools and equipment in removing water from floors (4); physical ability to work from ladders (19); ability to crawl on hands and knees (29); ability to drag or carry adults or children (13); ability to push or lift heavy objects (14); manual dexterity (3); physical ability to perform first aid and cardiopulmonary resuscitation (15); physical ability to assist in loading of hose bed (2); physical ability to participate in physical training (2).

9

In meeting with firefighters in this context, Landy identified himself as someone hired by the City to help defend this lawsuit, Tr. 1001, which was criticized at trail by Dr. Cranny. Tr. 359-361. This approach strikes the Court as ill-advised and unnecessary.

10

These problems are from three potential sources of bias in the correlations: restriction of range, measurement errors in the criteria and the use of dual eligible lists for hiring. These are discussed in Jt. Ex. 16, at 5-4 to 5-16. Kriska/Hines resolved these problems in a highly conservative manner.

11

At trial, Dr. Landy testified regarding possible rescoring of the 1984 examination on a pass/fail basis. Passing levels were set at the mean score for all female applicants on the 1984 physical examination. Virtually no females scored above the mean, thus determined, on all five test events; on the other hand, a very large number of males scored above the means. Hence, Landy testified, a pass/fail examination would have even greater adverse impact than a scored examination. Tr. 1029-1032, The problem with this is that the pass points are simply arbitrary; there is no attempt to base them on a job analysis.

United States District Court

SOUTHERN DISTRICT OF OHIO

ANN BRUNET, et al., JUDGMENT IN A CIVIL CASE

V.

CASE NUMBER: C-2-84-1973

CITY OF COLUMBUS, et al.,

☐ **Jury Verdict.** This action came before the Court for a trial by jury. The issues have been tried and the jury has rendered its verdict.

☐ **Decision by Court.** This action came to trial or hearing before the Court. The issues have been tried or heard and a decision has been rendered.

IT IS ORDERED AND ADJUDGED that on the issue of liability JUDGEMENT is for the defendents on the §. 1983 claim and the Title VII claim pertaining to the 1980 examination, and JUDGMENT is for the plaintiffs on the Title VII claim pertaining to the 1984 examination.

May 14, 1986 Kenneth J. Murphy
Date Clerk

(By) Deputy Clerk

FILED May 30,1986

UNITED STATES DISTRICT COURT
SOUTHERN OHIO DISTRICT
EASTERN DIVISION

ANN BRUNET, et al.,
Plaintiffs,

v. Case No. C-2-84-1973

CITY OF COLUMBUS, et al.,
Defendants.

OPINION AND ORDER

This matter comes before the Court to consider the remedial decree to be entered in this case. In an Opinion and Order entered May 13, 1986, the Court concluded that the defendents discriminated against the plaintiffs--a class of past, present and future female applicants for firefighter--when they used the 1984 firefighter examination to select firefighters. In light of this conclusion, the Court directed the parties to file their suggestions regarding a remedial decree. Having considered these suggestions, the Court now

renders its decision regarding a remedy.

Under Title VII, hiring by use of an examination with adverse impact on women constitutes impermissible discrimination unless that examination has been proven to be job-related. The Court concluded that defendants had failed in two respects to show that the 1984 firefighters' test was job-related. First, the physical abilities measured by the physical test, taken as a whole, did not reflect the physical abilities actually used on the job. Second, defendants used test scores to rank-order applicants who were to be hired in order from eligibility lists. The practice of ranking is permissible only if there is evidence that scores on the test vary directly with job performance. The Court concluded that the defendants had failed to present such evidence. The Court stated that to

eliminate the problem of lack of representativeness of the physical test, defendants must redesign the test. The Court further stated that, to eliminate the problems associated with ranking, the defendants must produce the requisite evidence or, alternatively, eliminate ranking.

While this case was pending, the City administered a new firefighters' entrance examination. In December 1985, a new written examination was administered. And, in March 1986, after this case had been tried, but prior to the submission of briefs by the parties, a new physical test was administered. The new physical test included two new events; in addition, some of the events on the 1984 physical test were modified in various ways before they were administered to applicants in 1986. The defendants state that the grading and scoring of the 1986 firefighter's examination has been suspended

pending this Court's determination of the job-relatedness of the test and acceptance of a scoring method. They also state that, for the first time, the firefighters' examination was administered to incumbent firefighters. The defendants refused any discovery to plaintiffs regarding the nature of administration of the 1986 physical test. Thus, the court has before it only the assertions of defendant's counsel regarding the 1986 test.

In their submission to the Court, plaintiffs have made a number of suggestions regarding particular events that were included in the 1984 physical test. For the moment, these suggestions are moot, because defendants have already redesigned the physical test. Some of the changes proposed by plaintiffs have already been introduced. Plaintiffs also propose certain changes in the scoring of the test.

In particular, they propose that all timed test events and all ranking based on physical test scores be eliminated. They also propose that defendents be required to administer any new test to incumbent firefighters, a demand that has already been met by the defendants. Further, plaintiffs demand timely access to information regarding the new test. And, they request an interim award of attorney's fees.

In their submission to the Court, defendants set forth a remedial plan in some detail. First, to address the Court's concerns regarding representativeness of the test, defendants propose to employ the same strategy used in the Landy, Jacobs study, Jt. Ex. 17, to evaluate the content validity of the 1984 physical test. In the Landy report, incumbent firefighters were asked to evaluate the physical abilities used on the job. A panel of industrial psychologists assessed the physical abilities measured by the test. The judgments of the

firefighters were than compared with those of the industrial psychologists to determine the representativeness of the test. The Court relied upon the results of this analysis to conclude that the 1984 test did not accurately reflect the requirements of the job. Defendants propose to present information regarding the 1986 physical test to a panel of industrial psychologists. They believe that the modifications introduced in 1986 have produced a valid, job-related test. Defendants also suggest that the scores of particular test events can be reweighted, if necessary to ensure representativeness. Defendants estimate that such a study of the 1986 physical examination could be completed by August 1986.

To address the Court's concern regarding use of test scores to rank-order applicants, defendants state that they have been discussing with their expert, Dr. Landy, a criterion-related validity study to address

issues concerning the scoring of the examination. Because the 1986 physical examination was administered to incumbent firefighters, defendants propose a concurrent validity study to compare the job performance of these incumbents with their test performance. Also, defendants point out, test scores of incumbent firefighters can be used to calculate cut scores for administration of a pass/fail physical examination, it necessary. In addition, defendants express interest in conducting a predictive validity study. If they are permitted to hire on the basis of the 1986examination, defendants could follow the development of applicants who are hired, comparing test scores with training success and job performance. A predictive criterion-related validity study provides the best and most direct evidence of job-relatedness.

In light of the proposals of the parties and the present circumstances, the Court concludes that the following remedial plan would most directly eliminate the discrimination and protect the interests of the defendants in a safe and efficient fire division. Defendants will be permitted to hire on the basis of the 1986 examination only when the content validity of the modified examination has been proven. The Court considers the proposal to duplicate the Landy study of the 1984 examination to be a reasonable way of making this showing. If the 1986 examination is shown to be content valid--or, if it is not, some further modified examination has been shown to be content valid--defendants must notify all female applicants for firefighter in 1984 of the new examination and administer it to all such applicants who are still interested in being considered for the position of firefighter.

When the content valid test has been formulated, and administered to 1984 female applicants, defendants may hire on the basis of the 1986 examination. At this point, the 1986 examination may be used only on a pass/fail basis, where the cutpoints defining passing scores are determined by the performance of incumbents on the examination in accord with the standards of the Uniform Guidelines, 29 C.F.R. §. 1607.1 et seq. Persons achieving pass scores on the test as a whole will be available for hire; if there are more passing applicants than positions available, candidates shall be considered for hiring so that the percentages of females and males considered for hire reflects the relative proportions of male and female applicants achieving passing scores. Defendants may determine the particular method by which this result is achieved. Further, before hiring from the 1986 examination, defendants must determine

the number of females who would have been hired in 1984 and the 1984 examination included a content-valid physical test and the test as a whole had been scored on a pass/fail basis. To the extent that there are such females, a set-aside of places in the firefighter classes hired on the basis of the 1986 examination must be created. Females applicants in 1984 who have taken the new firefighter examination will then be considered to fill these set-aside positions in the order of their total test scores. When hiring from the 1986 list occurs, defendants will then perform their proposed predictive criterion-related validity study. If defendants can show, on the basis of the predictive and concurrent criterion-related validity studies they propose, that the test as a whole is sufficiently precise to be used for ranking, they may then apply to the Court for an Order permitting ranking.

The Court is satisfied that this remedy achieves a reasonable accommodation of plaintiffs' interest in freedom from discrimination and relief from any effects of past discrimination, and the defendants' strong interest in safe and efficient staffing of the division of fire. Through this remedial process, the Court believes that the central goal of Title VII--to make job qualifications the controlling factor and factors such as sex irrelevant--can best be realized. For these reasons, the Court makes the following order:

1) The defendants are enjoined from any hiring of entry-level firefighters on the basis of the 1986 firefighter examination until here has been compliance with this Order.

2) As soon as practicable, defendants shall submit to the Court a report detailing the results of an analysis of the 1986 physical test comparable to that underlying Appendix R of the Landy, Jacobs

report, Jt. Ex. 17. No members of the Landy, Jacobs firm or any employee of the City of Columbus may serve as a member of the panel of industrial psychologists that is to evaluate the test. Prior to the analysis of the1986 test by the expert panel, all relevant materials and the design of the study shall be made available to an expert to be chosen by the plaintiffs, who will review these materials and file a written report to accompany the City's proposed report. The fees of this expert will be paid by the defendants. Further, the defendants shall provide discovery regarding the content and administration of this study; plaintiffs shall file their objections, if any, contemporaneously with defendants' report. The Court shall then determine whether the 1986 physical examination is content valid. If it is not, the defendants must redesign the test and again demonstrate its content validity. If the test must be

remodified, it must again be administered to incumbent firefighters.

3) If the 1986 physical test if found to be content valid by the Court, the defendants shall then submit promptly to the Court a report on the results of the administration of the 1986 physical examination to incumbent firefighters. If the 1986 physical test already administered is found not to be valid, the defendants shall then remodify the test, and, when the test has been determined to be content valid, shall administer the test to incumbent firefighters and report on the results. The report shall describe fully the details of administration of the test to incumbents and provide data regarding the incumbents. The report shall include a proposal regarding cutoff scores to be used to grade the test as a whole pass/fail. These cut-off scores shall "be set so as to be reasonable and consistent with normal expectations of acceptable

proficiency within the work force." 29 C.F.R. §. 1607(H). Defendants shall provide discovery to plaintiffs regarding the test of incumbents and all relevant information regarding characteristics of incumbents. The parties shall present to the Court their proposals regarding appropriate cutoff scores. The City shall show the number of males and females that would be hired under each proposal. The Court will then determine appropriate scoring procedures.

4) When defendants have formulated a content valid physical test and pass/fail scoring procedures have been determined by the Court, defendants shall notify all female applicants for firefighter in 1984 that they may reapply to take the firefighter examination as a result of this Court's decision. The form of notice must be approved by the Court. Only female applicants on the 1984 eligibility list who were considered for hire and rejected for reasons

other than the physical test are excepted from receiving notice. Defendants shall then administer the new physical test to all applicants who appear in response to the notice. Defendants shall devise a training program for applicants, and shall notify all 1984 applicants of the availability of this program. Plaintiffs shall make their suggestions to defendants regarding the content of this training program in writing in a timely manner. The test results will then be scored in the manner previously approved by the Court; defendants may use 1984 written test scores to determine the total score of female applicants who retake the physical test. Defendants shall determine the number of female applicants who would have been hired in 1984 had a content valid test been administered and had the test as a while been graded pass/fail. For this purpose, the defendants shall assume that, if a greater number of male and female applicants achieve passing scores than were in fact

hired from the 1984 eligibility lists, male and female applicants would have been considered for hiring in proportion to the relative proportions of male and female applicants achieving passing scores.

6) When defendants have completed these steps, they may hire on the basis of the 1986 examination. Defendants shall set aside the number of places determined as detailed in §. 5 of this Order to represent female applicants who would have been hired from the 1984 eligibility lists, and shall fill these positions, if any, before any other hiring. In addition, defendants shall hire males and females in proportion to the relative proportions of males and females achieving passing scores.

7) Plaintiffs have prevailed in part upon an issue determining the rights of the parties, and, thus, an interim award of attorney's fees and costs is

appropriate. Plaintiffs shall make application for interim fees, and, after considering defendants response, the Court will render its decision.

8) Defendants shall disregard any local ordinance, law of Ohio, charter provision or Ohio constitutional provision to the extent that it conflicts with the implementation of this Order.

IT IS SO ORDERED

_____

UNITED STATES DISTRICT JUDGE

United States District Court

SOUTHERN DISTRICT OF OHIO

ANN BRUNET, et al., JUDGMENT IN A CIVIL CASE

V.

CASE NUMBER: C-2-84-1973

CITY OF COLUMBUS, et al.,

[ ] **Jury Verdict.** This action came before the Court for a trial by jury. The issues have been tried and the jury has rendered its verdict.

[ ] **Decision by Court.** This action came to trial or hearing before the Court. The issues have been tried or heard and a decision has been rendered.

IT IS ORDERED AND ADJUDGED the parties are to comply with the terms and conditions of the Opinion and Order entered May 30, 1986. Said order of judgment shall be dated nunc pro tunc as of May 30, 1986.

nunc pro tunc
May 30, 1986 Kenneth J. Murphy
Date Clerk

(By) Deputy Clerk

FILED May 21, 1987

UNITED STATES DISTRICT COURT
SOUTHERN OHIO DISTRICT
EASTERN DIVISION

ANN BRUNET, et al.,
Plaintiffs,

v. Case No. C-2-84-1973

CITY OF COLUMBUS, et al.,
Defendants.

OPINION AND ORDER

This matter is before the Court to consider the content validity of the firefighter physical capability test developed and administered by the City of Columbus ("the City") in 1986. Before addressing the merits of this matter, a brief review of the history of this case is in order.

The plaintiffs originally brought this case under Title VII of the Civil Rights Act of 1964, 29 U.S.C. §. 2000 et seq., challenging certain parts of the tests used by the City in 1980 and 1984 to select entry-level

practice of rank-order hiring was impermissible under Title VII and the Uniform Guidelines since the defendants' could not establish that higher scores on the test varied sufficiently with job performance to justify rank-order hiring. Id. at 1248-1249.

In order to remedy the defendants' violation of Title VII, the Court prohibited the City from hiring any new firefighters until it could validate a new physical capacity test which the City had developed and subsequently administered in early 1986. Id. at 1253. The Court also prohibited the defendants from rank-order hiring firefighters until they could complete a criterion-related validity study and demonstrate that small differences in test scores were related to job performance. Id. at 1252-1253. Until they did so, the defendants were required to establish a pass/fail scoring system for the physical capability test based upon the scores of incumbent firefighters and to hire

men and women in proportion to the number passing the 1986 exam. Id. at 1253. Finally, upon adoption of a content valid test, the defendants were required to offer the new test to the 1984 female applicants and hire the number of said women who would have been hired in 1984 had the earlier test been content valid. Id. at 1253.

However, in the interim period, the defendants, citing the need to hire new firefighters, moved for a stay of the Court's injunction and requested that they be able to hire an interim class of firefighters. The Court granted the defendants' motion, but required the City to establish a cut-off score to be used in conjunction with interim hiring and to hire men and women for the interim class on the basis of the proportion of men and women passing the

1986 test. Id. at 1256. Thereafter, the Court approved the use of a cut-off score for the physical capability test set at one standard deviation below the mean incumbent score for the physical test graded as a whole. See Opinion and Order dated October 8, 1986 at 5-6. As a result of this cut-score, 13 out of 63 females and 505 out of 853 males passed the 1986 test. See Joint Proposed Scoring Scheme for Interim Hiring at 1-2.Pursuant to the Court's order requiring proportional hiring, the City hired one woman from the 1984 female firefighter candidate population and two women from the 1986 firefighter candidate population for the first interim class. Opinion and Order dated November 20, 1986 at 2-3.

In March 1987, this Court held a three-day hearing to consider the issue of whether the defendants' 1986 physical capability test is

content valid. In Part I of this Opinion, the Court describes the 1986 physical test and the study conducted by the City to establish its validity. Thereafter, the Court discusses the exam and concludes that the 1986 exam as administered by the City is not content valid because of the inclusion of a non-critical task on the test which has an adverse impact upon women. However, elimination of this event results in an exam which meets the requirements of Title VII. Accordingly, the Court concludes that a modified version of the 1986 physical capacity examination is content valid. At the hearing in March 1987, the Court also heard evidence introduced by the plaintiffs concerning alleged irregularities in the administration of the physical capability test to incumbents and applicants. Although the Court finds in Part II of this Opinion that some irregularities did occur in the administration of the test to

incumbent firefighters, these deficiencies were not of a sufficient nature to render the incumbent scores unreliable or inaccurate for the purpose of using them as a basis for cutoff scores.[1]

Finally, in Part III of this Opinion, the Court considers and discusses evidence concerning the cutoff score that should be used in conjunction with evaluating the 1986 physical exam. As in its Opinion and Order dated October 8, 1986, the Court rejects defendants' proposal to utilize a multi-event cutoff score method and, instead, adopts plaintiffs' proposal to base the cutoff score on the physical test graded as a whole. However, the Court accepts the defendants' proposal to set the cutoff at a score equivalent to one standard deviation below the incumbent mean score.

Part I

As early as 1984, the City began to work on the development of the 1986 physical capability exam. The events chosen for the 1986 physical exam were asked on job analyses performed by Julia Ingram and David Kriska[2] in 1980 and the Landy-Jacobs consulting firm in 1985, as well as on the experiences of other municipalities. Kriska Test. After arriving at the events to be used on the 1986 physical exam, the City proceeded to conduct structured interviews with nineteen firefighters in order to "fine tune" the events. Id. and Def. Ex. M. As a result of this process, the City arrived at a physical capability firefighter test comprised of seven events, most of which were different from the seven events tested by the City on the 1984 physical capability exam. The test was then administered to incumbent firefighters. Kriska Test.

Based on the experience of the incumbents, several events and the order in which they were given were modified. Id. Thereafter, the test was administered to the 1986 firefighter applicants. Finally, portions of the test were re-administered to the incumbents in order to generate a complete incumbent firefighter data set. Id.

The events comprising the 1986 physical capability exam and the scores received by the incumbent firefighters and 1986 applicants were as follows:

1) Stairway Climb: This event entailed shuttling several types of equipment up two flights of stairs in four separate trips while wearing a turnout coat and an air tank. The equipment included a hose, fire extinguisher, air bottle, and an exhaust fan. Id.; Ex. M. The stair climb was a timed event intended to simulate the method used by the Fire Division in

shuttling equipment in a high-rise fire. Kriska Test. Timing of the event was completed when the last piece of equipment was carried up to the second floor. Incumbent firefighters were told to perform the event at the same pace they would at a fire. Id. Applicants were told that the event was timed and, therefore, that faster speeds would result in better scores. However, they were also warned to pace themselves in order to avoid becoming exhausted before the event was completed. Def. Ex. S. After placing the last piece of equipment on the top level, applicants were required to return the equipment to the first floor, carrying a separate piece of equipment each trip. Applicants were required to return all of the equipment within six minutes. If an applicant had any time remaining, he or she could rest for the balance. However, if an applicant took more than six minutes to return the equipment, additional time was added on to his or her

score. Ex. M.

Unlike the stairway event on the 1984 test, the 1986 stairway event more closely simulated actual firefighter tasks. Brunet I at 1233. Moreover, the 1986 stairway event was a much better measure of stamina than the 1984 stairway event since it required over three minutes on average to complete, whereas the 1984 stairway event required slightly more than one minute to complete. Id. at 1233.

The mean incumbent score for the stairway event was 183 seconds while the means scores for male and female applicants were 144 seconds and 239 seconds, respectively. Ex. 11; Ex. I. The standard deviation of the incumbents' score was 47 seconds meaning that two-thirds of the incumbents completed the event in a time ranging between 136 and 230 seconds. Thirty-four female applicants received scores within this range.[3] Ex. G.

2) **Search Event:** The second event on the test was a simulated search event. This event has no parallel on the 1984 event. It was designed to simulate the firefighter's task of feeling his or her way through a smoke-filled room. Applicants were required to crawl through a 75 foot long maze while wearing a turnout coat, air tank, helmet, and blackened goggles. Inside the maze were two 80 lb sand bags which simulated obstacles which a firefighter might encounter during a search. Like the stairway event, this event was also timed.

The mean incumbent score for this event was 30.35 seconds with a standard deviation of 11.6 seconds. Def. Ex. I. The mean scores for the male and female 1986 applicants were 22.57 and 32.13 seconds, respectively. Ex. 11. Thus, the average female applicant performed this event in roughly the same time as the average incumbent firefighter. In fact, fifty-two of the female applicants performed this

event in better time than the bottom fifteen percent of incumbent firefighters. Ex. G.

3) **Rescue Event:** This event was similar to the carry/drag event that was on the 1984 firefighter exam except several important improvements were made. Like the 1984 carry/drag event, this event required an applicant to manuever a 125 pound sandbag/dummy through a 35-foot serpentine obstacle course. However, unlike the 1984 event, where applicants were required to carry the sandbag and were penalized if they dropped or dragged it, the 1986 rescue event required applicants to drag the dummy through the course while keeping at least one knee on the ground with each step. Ex. M; Kriska Test. Thus, this event comes much closer to simulating the actual fire situation where the smoke and heat of a fire typically require firefighters to keep victims close to

the ground and, hence, require them to drag the victim out of a building. Brunet I at 1232. Applicants were required to complete this event while wearing a turnout coat, air tank, regulator, and harness. The only criticism the Court has of this event is that the City could have used an articulated dummy to better simulate actual conditions.

As with the 1984 carry/drag event, the 1986 rescue event was also timed. The mean incumbent score was 27.28 seconds with a standard deviation of 11.6 seconds. Ex. I The mean score for the male and female 1986 applicants were 17.13 and 36.28, respectively. Ex. 11. Thus, on average, female applicants did significantly worse on this event, a fact which is not surprising given this event's emphasis on anaerobic capacity and upper-body strength. However, notwithstanding this fact, of the sixty-two female applicants who took the 1986 physical test,

nineteen received scores better than the average incumbent firefighter score and a total of thirty-seven female applicants performed the event in better time than the bottom fifteen percent of incumbent firefighters. Ex. G.

The Court notes that the plaintiffs criticize both the search and rescue events as being unrepresentative of the actual tasks. They further maintain that there is no evidence that all-out speed is required in performing these tasks. The Court finds no merit to either of these contentions. Although the Court has little doubt that improvements could be made to both events to make them more realistic, under the circumstances, the Court finds both events to be reasonable approximations of the work and abilities required of firefighters in performing these events. Moreover, the overwhelming weight of the evidence in this case clearly indicates that

speed is of critical importance in performing these two tasks. Questions 2 and 3 to Ex. 6. The Court further notes that both tasks are judged by firefighters as being relatively important, Appendix F to Ex. K, and measure abilities which firefighters deem to be critical. Brunet I at 1239, n. 7.[4]

4) Coupling Event: This event was virtually identical to the coupling event on the 1984 test. Applicants were required to screw three metal plugs into three threaded intakes on a piece of fire equipment, a multiversal, and then unscrew them. The event was timed and a faster time resulted in a higher score. The only difference between the 1984 and 1986 tests was that applicants were given two opportunities to attempt the event and were given the better of the two scores. The mean incumbent score was 22.84 seconds with a standard deviation of 4.54 seconds. Ex. I. The mean scores for the male and

female 1986 applicants were 21.80 and 25.54, respectively. Ex. 11. Forty-one female applicants performed the event in better time than the bottom fifteen percent of the incumbent firefighters. Ex. G.

5) **Hose Hoist:** This event was also similar to the equipment hoist event on the 1984 test. This event required applicants to hoist 100 feet of uncharged three-inch, plastic hose to the first level of a parking garage (42 feet). Applicants were required to first hoist the hose with rope. Once the hose reached the first level, applicants were required to pull the rest of the hose up to the first level. Applicants had the option of using a hose-roller to hoist the hose or to just pull it over the edge of a railing. Ex. M.

This event was intended to simulate hoisting hose or equipment that would be needed at a fire in the upper story of a multiple building. Id.

The plaintiffs contend, however, that this event neither represents a critical job task nor approximates the actual job task. The Court agrees on both points. As revealed by the Ingram job analysis, this task is relatively unimportant. Appendix 5 to Ingram Job Analysis;[5] Brunet I at 1232-1233. It also appears to be infrequently performed. Indeed, seven of the nineteen firefighters interviewed by the defendants could not remember ever performing this task. See Ex. 6 at Question 1. With respect to the second point, a review of the firefighter interviews indicates that only two firefighters ever hoisted uncharged hose. More typically, individual firefighters would hoist smaller items of equipment or groups of firefighters would hoist heavier items, charged hose, or even persons. Id. at Question 1(a) (2), (3), and 1(b). Moreover, when hoisting was done, more often than not it was done at a paced rate, not at an all-out speed as required in the

1986 test. Id. at Question 1.

As with the prior events, this event was also timed. The mean incumbent score was 58.65, with a standard deviation of 21.44 seconds. The mean scores for the male and female 1986 applicants were 63.11 and 144.81 seconds, respectively. Like the rescue event, this event emphasized primarily muscular strength and anaerobic capacity of the arms. Brunet I at 1233. Consequently, women did significantly worse on it than men. Only three female applicants received scores better than the average incumbent firefighter, while only a total of four women did better than the bottom fifteen percent of incumbent firefighters. Ex. G.

6) **Pike Pole push/pull:** This event was designed to simulate the use of a pike pole to tear down ceilings and walls to look for smoldering fires.

It was basically a modification of a similar event on the 1984 test. While wearing a turnout coat, air tank, regulator, and harness, applicants were required to push up on a pike pole handle attached to a sixty-pound weight until it touched a stop plate two times. An applicant would then switch to a different handle which was attached to an eighty pound weight and pull it down until it touched the floor one time. Applicants were scored according to how many times they could repeat this sequence in a four minute time period.

As with the hose hoist, the plaintiffs claim that the pike pole event represents an unimportant job task. The Court disagrees with this contention. In both the Ingram and Landy job analyses, firefighters identified the ability to use a pike pole and the task of overhaul as being important. Appendix H to Ex. K; Ingram Job Analysis, Appendix 5.

The plaintiffs also contend that the event does not approximate the actual job task. More specifically, they argue that the event over-emphasizes speed, has the wrong-ratio of push-to-pull, lasts too long, and is an awkward event. The Court finds little merit to any of these contentions. After reviewing the firefighter interviews, the Court finds that the ratio of push-to-pull of 2:1, the weights involved, and the duration of the event are reasonable approximations of the actual task. Kriska Test.; Question 4 to Exhibit 6. the Court also finds the design of the event to be reasonable, particularly considering the fact that the City was attempting to design an event which came closer to simulating the strength required in the separate push and pull components of using a pike pole. The Court agrees with the plaintiffs, however, that the test over-emphasizes speed. There is no

evidence in this record to indicate that speed is of the essence in operating a pike pole. Indeed, given the fact that the pike pole is used primarily in over-haul, it is more likely that firefighters use the pike pole at an even pace. Brunet I at 1232.

The average number of repetitions performed by incumbents on this event was 46.76, with a standard deviation of 10.0. Ex. I. The mean number of repetitions by men and women applicants in 1986 were 41.53 and 27.97, respectively. As with the hose hoist, there is a significant difference in the average number of repetitions performed between incumbent firefighters and female applicants. Only two female applicants completed more repetitions than the average incumbent firefighter and a total of six female applicants did better than the bottom fifteen percent of incumbents.

7) **Fan Hang.** The final event on the 198 physical test was the fan hang event. Applicants wearing the same break-out gear as in the pike pol event, were required to hang a bracketed hook on the top of a standard-size door, and then hang an exhaust fan weighing approximately forty pounds on the hook. Applicants had to complete this event in twenty-one seconds. This cutoff was based on the greatest amount of time an incumbent firefighter took to complete the event. The event was scored on a pass/fail basis. Ninety-nine percent of the male applicants and eighty-one percent of the female applicants passed this event.

In addition to the above changes in the event, the manner in which the 1986 test was administered was also changed. In 1984, applicants were tested in groups. Consequently,they would frequently have ten

minute rests between events. This contributed to the overall anaerobic nature of the 1984 test. However, in 1986, applicants were tested individually and were accompanied throughout the test by a monitor. Thus, an applicant would complete an event and, except for the stairway climb where a rest period was built into the procedure, would immediately move on to the next event. Because of the physical lay-out of the test, an applicant was required to walk several hundred yards between the stairway climb and search events, the coupling and the hose hoist events, and, finally, between the hose hoist and pike pole events. Landy Test. Once arriving at each event, an applicant would take a minute or two to put on equipment, if necessary, and listen to instructions about the event. Thus, the test as a whole was administered in a fashion which minimized the time periods which applicants could rest and recover from the previous events and,

therefore, improved the test as a measure of stamina.

In order to demonstrate the content validity o the above test, the City followed a methodology whic was developed by the Landy-Jacobs consulting firm an approved by the Court in its earlier Order as bein consistent with the procedures outlined by the Uniform Guidelines. Brunet I at 1248, 1253. More specifically the Court directed the defendants to perform a conten validation study similar to the one performed by Landy-Jacobs for the 1984 test. Id. at 1253.

In their earlier content validity study, Landy and Jacobs conducted a job analysis wherein they divided the job of firefighting into various task groupings. Incumbent firefighters were asked to rate the importance of these various tasks and to evaluate the physical abilities that were necessary to

perform the various job tasks. Id. at 1240-1241, 1244. This latter part of the job analysis was done by instructing firefighters in the Fleishman taxonomic classification system and then asking the firefighters to distribute one hundred points across the various Fleishman abilities to reflect their relative role in the performance of a particular group of tasks. Id. at 1240. The scores assigned to each ability were then averaged across the various task areas, each appropriately weighted according to importance of the task, to arrive at an estimate of the relative importance of each ability for the job of firefighting. Id. at 1240.

Dr. Landy next asked a group of independent industrial and organizational psychologists to make an evaluation of the 1986 test similar to that made by the firefighters. As with the firefighters, the industrial psychologists were provided with definitions of

the Fleishman abilities. They were also give information about the 1986 test and were shown videotape of the test in its entirety. Finally they wer allowed to ask any questions about the test. Landy Test.

For each event, the psychologists were asked to distribute one hundred points across the various abilities to reflect the extent to which each ability was tested by the particular event. These results were then averaged across the various events to yield an overall measure of the extent to which a given ability was measured by the 1986 physical examination. They were also asked to rate the test as a whole instead of one event at a time. Id. Dr. Landy pointed out that the intra-class correlation coefficients of the judgments by the industrial psychologists concerning the abilities measured by each event was very high, ranging between .92 and .99. Ex. L. Thus, he concluded that

the industrial psychologists' judgments were highly reliable. Landy Test.

The eight highest rated abilities by the industrial psychologists as averaged across all seven events as compared with the ratings of the same abilities by the firefighters are as follows:

Ability	Score*	Rank in Firefighters' Rating*	Score in* Firefighters' Rating*
Dynamic Strength	6.19	4	4.81
Stamina	5.35	1	8.17
Static Strength	5.04	2	8.11
Dynamic Flexibility	4.85	12	1.93

Gross Body Coordination	4.58	7	2.63
Explosive Strength	4.56	3	4.86
Speed of Limb Movement	4.52	14	1.19
Multilimb Coordination	4.06	5	.92

*Source: Ex. B

Qualitatively, the Court finds the match between the 1986 test and the abilities required to perform the job of firefighter to be better than the match with the 1984 test. Brunet I at 1244. For example, in the 1984 test, speed of limb movement and dynamic flexibility were the two top abilities and accounted for over thirty-two percent of the points awarded by the industrial psychologists. In contrast, the firefighters has ranked these two abilities as

thirteenth and fourteenth in relative importance, accounting for only eight percent of the total points awarded. Ex. B. However, these two events are ranked lower (fourth and seventh) and account for a lower portion of the total points (eighteen percent) on the 1986 test. Moreover, abilities ranked relatively high by the firefighters, such as stamina, static strength, explosive strength, and dynamic strength, are represented to a greater extent and ranked higher on the 1986 test than on the 1984 test.

On a more quantitative basis, the defendants point out that the statistical correlation between the judgments of the firefighters and the industrial psychologists is greater for the 1986 test than the 1984 test. Specifically, Dr. Landy calculated that the correlation coefficient between the firefighters' judgments and the psychologists' judgments

averaged over all seven events is .65. Dr. Landy pointed out that this correlation is statistically significant at the .003 level and, furthermore, that it is substantially higher than the correlation coefficient for the 1984 test, which was .44. Landy Test.; Ex. L; Brunet I at 1245. Dr. Landy further testified that the correlation coefficient between the firefighters' judgments and the industrial psychologist judgments' for the test viewed as a whole is .78, which is statistically significant at the .0001 level. Ex. L; Landy Test. Based on the improvement in the 1986 test and the substantial relationship between the ratings of the industrial psychologists and firefighters as measured by the above correlation coefficients, Dr. Landy concluded that the 1986 test is content valid. Landy Test; Ex. L.

The plaintiffs raise a variety of objections to and criticisms of the methodology used to

validate the test and the conclusions reached by Dr. Landy. They first argue that the City developed the test based upon an unrepresentative sample of nineteen firefighters. However, the Court finds that the City relied upon the various job analyses it conducted as well as the judgment of professional industrial psychologists in developing the 1986 test. The interviews it conducted with the nineteen firefighters were only for the purpose of further refining the specific events. Kriska Test. Although it is true that the defendants should have "fine tuned" these events based upon a representative sample, Cranny Test., the Court is satisfied that the information derived from these interviews and other evidence presented in this proceeding provides a sufficient basis for designing these events, as well as for evaluating whether the events approximate actual job tasks as required by the Uniform Guidelines.

The plaintiffs next express several objections concerning the reliability of the psychologists' judgments. First, they contend that the psychologists were not provided with proper operational definitions of the abilities involved in the work behavior observed. More specifically, they point out that the operational definitions of the Fleishman taxonomic abilities provided to the psychologists were different from those provided to the firefighters in 1985 when Dr. Landy conducted his job analysis. These differences primarily involved the addition of examples to the definition of a particular ability. Particularly troubling are those additions which involved a test event as an example of the taxonomic ability in question. For example, in defining speed of limb movement, the defendants added an example of tearing down a ceiling with

a pike pole.[6] The plaintiffs argue that these changes cued the psychologists and biased their judgments concerning abilities measured by a particular task.

The plaintiffs raise a serious issue. However, although the Court believes that the defendants departed from accepted professional scientific methodology by supplementing the Fleishman taxonomic definitions with examples of actual test events, Cranny Test., the Court does not believe that such additions skewed or biased the judgment of the psychologists. The Court arrives at this conclusion after reviewing the psychologists' judgments for each of the test events in question. Again, using the same example as above, even though the defendants added an example of using a pike pole in the definition of speed of limb movement, only four of the eighteen industrial psychologists allotted more than twenty percent of their points to this ability. A similar pattern

holds up for the other abilities in question, except for stamina.[7].

With respect to stamina, which included the stairway climb as an example, all but three of the industrial psychologists allocated more than twenty percent of their points to stamina. The Court does not find that this indicates a bias, however, because based on the prior testimony about aerobic capacity, it believes that the psychologists fairly estimated or perhaps even under-estimated the stamina component of the stairway test event. Therefore, although the Court agrees with the plaintiffs that the defendants should have provided the industrial psychologists with the same Fleishman taxonomic definitions it provided the firefighters, the Court finds that the defendants' departure from standard scientific methodology did not bias the psychologists' judgments nor render their judgments unreliable.

The plaintiffs next argue that there is too much variation between the industrial psychologists' judgments concerning the abilities measured by the 1986 test events to give them much them much credence. They discount defendants' relatively high intra-class correlation coefficients as being inflated because of the inclusion of a large number of irrelevant abilities in the calculation of the inter-class correlation coefficients.[8] While the Court appreciates that such inflation probably exists, it nonetheless finds, after its own review of the industrial psychologists' judgments, insufficient disagreement between said judgments to indicate that they are unreliable. Indeed, the fact that there is a significant degree of qualitative agreement between the judgments of these independant industrial psychologists' coupled with the fact that they ignored the potential cues discussed

above are highly probative of the reliability of their judgments.

The plaintiffs also argue that the definitions provided to the psychologists were incomplete since the psychologists were not instructed that test events which involve stamina must last more than a minute. In support of their argument, the plaintiffs point out that many psychologists rated test events which took one minute or less as having considerable amounts of stamina. They direct the Court's attention to the psychologists' ratings for the search, rescue, and hose hoist events. The Court has reviewed these ratings and finds plaintiffs' contention to be without merit. With but a few exceptions, the individual psychologists rated the stamina component of these events as being non-existent or miniscule. See Ex. U.

The plaintiffs next attempt to challenge the reliability of the psychologists by pointing to the differences in the 1985 and 1986 psychologists' ratings. They argue that these differences indicate that the psychologists were confused or at least guessing when rating the abilities measured by the physical exam. The Court declines to draw this inference. Even accepting that such a comparison would be analytically proper, the Court sees little evidence that the panels of independent psychologists who reviewed the 1986 test were confused or guessing when they evaluated the test. As the Court discussed above, there appears to be a substantial degree of correspondence between the judgments of theindividual psychologists who evaluated the 1986 test.

Finally, the plaintiffs challenge the reliability of the psychologists' judgment concerning the abilities

measured by the test overall. As the Court described above, the defendants' asked the psychologists to consider the test as a whole and rate the various abilities tested. In doing so, the psychologists took into consideration the time and activities between test events. The psychologists' resulting judgment yielded a much closer "match" with the firefighters' ratings, both qualitatively and as measured by the correlation coefficient.[9]

The plaintiffs argue that the psychologists' judgment of the test taken as a whole is unreliable because one cannot be sure of what the ratings are based upon. Indeed, this is the very reason that the defendants separated the job of firefighting into discrete tasks and asked the firefighters to rate the abilities required for each task instead of asking firefighters to rate the job as a whole.

Ex. K at 13. Although the Court believes that expert industrial psychologists are undoubtedly more competent than firefighters in keeping the various tasks in mind and rating them overall according to the Fleishman taxonomic abilities, the Court shares the plaintiffs' concern that there is no way to know which events the psychologists focused upon or whether they weighed the events disproportionately. Indeed, upon review of the psychologists' judgments, it would appear that the psychologists placed undue emphasis on what happened between events. Moreover, the Court observes that there was no attempt to standardize or otherwise assure that each applicant spent only a certain amount of time and was engaged in the same activities between events. Given these considerations, the Court finds that the correlation coefficient between the psychologists' ratings of the test overall

and the firefighters' ratings is too unreliable to be used as a quantitative measure of the relationship between these two sets of judgments.

The plaintiffs next challenge various aspects of the statistical analysis performed by the defendants. First, they argue that reliance on the statistical significance of the correlation coefficient between the ratings of the firefighters and psychologists is inappropriate and that such statistical significance is irrelevant. Cranny Test. Dr. Cranny argued at the hearing that a test of statistical significance only determines whether or not a particular sample is representative of a population of values, not whether the relationship between two sets of independent variables is statistically significant. Id.

The Court will be the first to admit that it possesses no great understanding of statistics and the use of t-tests to determine whether a particular

statistical difference or similarity is statistically significant. However, the Court agrees with Dr. Cranny that a test of statistical significance relates to whether a sample correlation coefficient is agood estimate of the population correlation coefficient. See D.W. Barnes, Statistics as Proof, 270-271 (1983); Wilson, Griggs and Job Testing, 58 Va. L. Rev. 844, 866-867 (1972). Thus, whether a positive correlation between two independent variables is "sufficiently high to influence the [C]ourt is not a statistical question." Statistics at 268-269. Rather, the question is whether the relationship between the two variables is practically significant. Id. at 269; Job Testing at 867; Note, Application of the EEOC Guidelines to Employment Test Validation: A Uniform Standard for Both Public and Private Employers, 41 George Washington L. Rev. 505, 522 (1973). Accordingly, the Court finds plaintiffs' objection to any reliance

on the statistical significance of the correlation coefficients to be well taken.

The plaintiffs further contend that the defendants' calculation of the correlation coefficients included irrelevant abilities and, therefore, inflated the true agreement between the ratings of the psychologists and thefirefighters. Plaintiffs' expert, Dr. Cranny, provided the Court with a straight-forward example of how, by adding a number of irrelevant abilities, to a statistical analysis, the degree of agreement between the psychologists' and the firefighters' judgments would increase required to demonstrate a significant relationship between the two sets of judgments. First, numerous courts have accepted lower correlation coefficients as being practically significant. See B. L. Schlei and P. Grossman, Employment Discrimination Law 129

(1983).[10] Second, the Court questions Dr. Cranny's logical basis in arriving at his opinion. As Dr. Cranny testified, a correlation coefficient is a relatively straight-forward descriptive statistic which serves as an index of the agreement between two variables or sets of numbers. Cranny Test. As the Court understands statistical analyses, however, a correlation coefficient provides no measure of how much variance in one variable is explained by another. Rather, one must calculate a regression coefficient. See Statistics at 294-315. Thus, the correlation coefficients calculated in this case provide no indication of how much of the differences in the firefighters' ratings are explained by the psychologists' ratings.

Finally, the plaintiffs argue that the 1986 test is not content valid because, like the 1984 test, it still over-emphasizes speed and dynamic flexibility and

under-emphasizes stamina. There is no question that dynamic flexibility and speed of limb movement are somewhat over-emphasized on the 1986 test. However, the extent of over-emphasis is greatly reduced from the 1984 test. As discussed above, speed of limb movement and dynamic flexibility were ranked as the top two abilities on the 1984 test. However, on the 1986 test, while these abilities still are ranked in the top seven by the psychologists, their relative importance decreased to fourth and seventh. More importantly, stamina is rated as the second most important ability measured on the 1986 test.

In considering the evidence, the Court notes that the over-emphasis of these abilities and the relative under-emphasis of stamina is partly due to the inclusion of the hose hoist event on the test. As the Court noted earlier, the task represented by this event

occurs infrequently, is considered to be relatively unimportant by firefighters, and fails to approximate the actual task involved. See supra at 10. Consequently, it is arguable that this event should not be included in the 1986 physical test. In fact, exclusion of this event results in a better qualitative and quantitative correspondence between the judgments of the firefighters and the psychologists. The eight highest rated abilities by the psychologists when averaged across six test events, excluding the hose hoist, as compared with the ratings of the same abilities by the firefighters are as follows:

Ability	Score* Rank in Firefighters'* Rating	Score in* Firefighters' Rating
Stamina	5.61 1	8.17
Static Strength	5.58 2	8.11
Dynamic Strength	5.56 4	4.86

Gross Body Coordination	5.07	7	4.86
Explosive Strength	4.87	3	2.63
Speed of Limb Movement	4.48	14	1.19
Wrist Finger Speed	3.95	15	.92
Multilimb Coordination	3.66	5	2.74

*Source: Exhibit 5 to Def. Ex. L. [11]

A comparison of the above table with the table on page 16, supra, clearly reveals that those abilities ranked highest by the firefighters generally have a higher value and are ranked comparatively higher by the psychologists when the hose event is deleted from their evaluations. Indeed, a comparison between the psychologists' ratings of the test with and without the

hose hoist event and the ratings of the firefighters reveals that the ratings of the test with the hose hoist deleted provides a better all-around match with the firefighters' ratings for eleven of the fifteen abilities.[12] With particular reference to those abilities which the plaintiffs maintain are either over-emphasized or under-emphasized, the Court observes that the importance of dynamic flexibility is significantly reduced and the importance of speed of limb movement is somewhat reduced when the hose hoist is deleted from the 1986 test events. Likewise, the importance of stamina is increased to the point that it ranks as the most important ability.Quantitatively, deletion of the hose hoist event also results in a higher correlation coefficient between the psychologists' and firefighters" judgments of .61. Of course, just as in Brunet I, the controlling question

is whether the degree of correspondence between the tasks and abilities measured by the 1986 test is sufficiently close to the actual tasks and abilities of the job of firefighting to establish that the 1986 test is content valid. This issue is addressed in the following section.

Legal Discussion

As the Court discussed in its first opinion in this matter, because the City's physical capacity examination has an adverse impact upon female applicants, it is the defendants' burden to show that the physical test "bears a manifest relationship to successful and efficient job performance." Harless v. Duck, 619 F. 2d 611, 616 (6th Cir.), cert. denied, 449 U.S. 872 (1980); Griggs v. Duke Power Co., 401 U.S. 424, 432 (1971). Under the Uniform Guidelines, 29 C.F.R. §. 1607.1 et seq., the defendants can meet their

burden by showing that the job-relatedness of a selection procedure is content valid. See Washington v. Davis, 426 U.S. 229 (1976); 29 C.F.R. §. 1607.14(C).

Under the Uniform Guidelines, in order to demonstrate the content validity of a selection procedure, a valid job analysis must be performed which identifies the important tasks required for the successful performance of the job. 29 C.F.R. §. 1607.14(C) (2). An employer must also show that:

> the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question.... In the case of a selection procedure measuring a knowledge, skill or ability, the knowledge, skill or ability being measured should be operationally defined... For any selection procedure measuring a ... skill, or ability the user should show that (a) the selection procedure measures and is a representative of that ... skill or ability; and (b) that ... skill, or ability is used in and is a

> necessary prerequisite to performance of critical or important work behavior(s). In addition, to be content valid, a selection procedure measuring a skill or ability should either closely approximate an observable work behavior, or its produce should closely approximate an observable work product. If a test purports to sample a work behavior or to provide a sample of a work product, the manner and setting of the selection procedure and its level and complexity should closely approximate the work situation.

29 C.F.R. §. 1607.14(C) (4).

Courts have interpreted the above section as requiring an employer to demonstrate that the content of the test is related to the content of the job and, furthermore, that the content of the test is representative of the content of the job.

Guardians Ass'n of New York City v. Civil Service Commission, 630 F. 2d 79, 95 (2d Cir. 1980), cert. denied, 452 U.S. 940 (1981); Vanguard Justice Society, Inc. v. Hughes, 592 F. Supp. 245, 258 (D. Md. 1984);

Stated another way:

> An examination has content validity if the content of the examination matches the content of the job. For a test to be content valid, the aptitudes and skills required for successful examination performance must be those aptitudes and skills required for successful job performance. It is essential that the examination test these attributes both in proportion to their relative importance on the job and at the level of difficulty demanded by the job.

Vulcan Society of New York City Fire Dept., Inc. v. Civil Service Commission, 360 F. Supp. 1265, 1274 (S.D.N.Y.), aff'd in part, remanded in part, 490 F. 2d 387 (2d Cir. 1973).

The Court notes that the representativeness

or proportionality requirement has been viewed with varying degrees of strictness. Several courts have held that a test must be shown "to examine all or substantially all the critical attributes [of a job] in proportion to their relative importance." Kirkland v. New York State Dept., 374 F. Supp. 1361, 1372 (S.D.N.Y. 1974), aff'd in part, rev'd in part, 520 F.2d 420 (1975), cert. denied, 429 U.S. 823 (1976); See also Firefighters Institute v. St. Louis, 616 F.2d 350, 359 (8th Cir. 1980); United States v. City of Chicago, 573 F.2d 416, 425-426 (7th Cir. 1978). Other courts, notably the Second Circuit, appear to have a less stringent standard. See Guardians, 630 F.2d at 98-99; Vanguard, 592 F. Supp. at 266.

In Guardians, the Second Circuit explained the importance of proportionality or representativeness:

> The reasons for a requirement that the content of the exam be representative is to prevent either the use of some minor aspect of the job as the basis for the selection procedure or the needless elimination of some significant part of the job's requirements from the selection process entirely; this adds a quantitative element to the qualitative requirement- that the content of the test be related to the content of the job. Thus, it is reasonable to insist that the test measure important aspects of the job, at least those for which appropriate measurement is feasible, but not that it measure all aspects, regardless of significance, in their exact proportions.

Guardians, 630 F.2d at 99. In the present case, to meet the representativeness requirement and establish that the 1986 test is content valid, the defendants must show that tasks tested by the events on the 1986 test are critical tasks which appropriately reflect those abilities which are required for the job of firefighter as determined by the defendants' job analyses.

The plaintiffs do not challenge the validity of the job analyses performed by the defendants. They do, however, object to the operational definitions used by the defendants. More specifically, the plaintiffs maintain that the definitions were incomplete and that the definitions provided to the psychologists had been changed from the definitions provided to the firefighters. As the Court discussed earlier, it finds little merit to these contentions.

The plaintiffs also contend that the 1986 test is not content valid because it fails to closely approximate observable work behaviors. 29 C.F.R. §. 1607.14(C)(4). For the most part, the Court concludes that the events tested on the 1986 test are fairly close approximations of the actual tasks which firefighters must perform on the job. The only exceptions are the hose hoist and the pike pole events. The hose hoist

required one person to lift an uncharged hose, whereas actual experience suggests that firefighters usually lift lighter items alone or hoist larger items in teams. With respect to the pike pole, it is clear that the event over-emphasizes speed by requiring applicants to do as many repetitions as possible in a four minute period. However, on the job, firefighters are much more likely to use the pike pole at a steady pace. Notwithstanding these deficiencies, which are far fewer than the deficiencies on the 1984 test, the Court concludes that the 1986 physical examination, judged overall, reasonably approximates the actual tasks performed by firefighters.

Finally, the plaintiffs contend that the defendants have failed to demonstrate that the content of the test reflects the content of the job or that the test is representative of the content of the job. More

specifically, the plaintiffs argue that the 1986 test includes events which do not represent critical or important tasks. In addition, they maintain that the 1986 test, like the 1984 test, over-emphasizes speed and dynamic flexibility and under-emphasizes speed and dynamic flexibility and under-emphasizes stamina to such an extent that it cannot be considered to be representative of the abilities required to do the job of firefighter.

Viewing the test overall, the Court believes that it is fairly representative of the abilities required to perform the job of firefighter. While it is true that some abilities are over-emphasized and other abilities are under-emphasized, in general, those abilities which the firefighters consider to be most important such as stamina, static strength, explosive strength, and dynamic strength, are rated by the psychologists as the

most important abilities measured by the 1986 test. Moreover, even though such abilities as speed of limb movement and dynamic flexibility are over--emphasized, their over-all importance is significantly diminished in contrast to the 1984 test. Finally, on a quantitative basis, the 1986 test is much more representative of the abilities required by the job as measured by the .55 correlation coefficient between the judgments of the firefighters and the psychologists.

However, viewing the individual test events, the Court cannot conclude that each task tested by the 1986 test is a critical or important task. As the Court discussed earlier, the hose hoist event is not considered an important task by firefighters. This fact is particularly troubling considering the significant impact this event has upon women applicants - only three women did better than the average

incumbent firefighter and only four did better than the bottom fifteen percent of incumbent firefighters. Because the cut score used in conjunction with this test is based upon the scores of incumbent firefighters, inclusion of this event on the 1986 test has a significant disparate impact on the number of female applicants who qualify for the job. Such an impact is quite permissible as long as the task involved is shown by a job analysis to be important, as is the case with the pike pole event.[12] However, where the event in question is unimportant, the result is a selection of job candidates on the basis of some minor aspect of the job. This is exactly the result which the representativeness requirement is intended to prevent. Guardians, 630 F.2d at 99. Consequently, it is a result which the Court believes is prohibited by Title VII. Therefore, the Court concludes that the 1986 test, as administered by the City, is not content valid.

However, by deleting the hose hoist even from the 1986 test, the Court believes that the 1986 test does fulfill the job-relatedness requirements of Title VII and the Uniform Guidelines promulgated thereto. Indeed, as the Court discussed above, see supra at 26, deletion of the hose hoist event from the 1986 exam results in a far better match or relationship between the abilities required for the job and the abilities measured by the exam. More to the point, deletion of the hose hoist event results in an improvement in the correlation coefficient between the firefighters' and psychologists' abilities ratings. Specifically, it places greater emphasis on stamina, while it de-emphasizes dynamic flexibility and speed of limb movement, thus substantially mitigating plaintiffs' criticism that the 1986 test is unrepresentative of abilities required to perform the job. Whereas, deletion of the hose

hoist event results in a test which selects firefighters only on the basis of important tasks and is more representative of the abilities required to be a firefighter, the Court hereby APPROVES the use of the 1986 physical capacity test minus the hose hoist event for testing firefighter applicants.

Part II

In addition to challenging the content validity of the 1986 test, the plaintiffs also challenge the circumstances and conditions under which the 1986 test was administered to incumbent firefighters. Although this issue is not relevant to the content validity of the 1986 test, it is critical in terms of establishing a cutoff score for the 1986 test based upon the performance of incumbent firefighters. If there were significant differences in the way in which the test was administered to incumbents and applicants,

the reliability of using the incumbents' test scores as a benchmark for establishing a cutoff score for applicants may be questionable. If such were the case, the incumbent scores could not be utilized to establish cut scores and incumbents would have to be re-tested.

The Court has already extensively described the 1986 test as it was administered to firefighter applicants. In a nutshell, each applicant took the entire test at one time and was accompanied by a monitor throughout the test event. In contrast, with the exception of one incumbent, all of the incumbents took some of the test events at separate times. In addition, the incumbents were tested in several different groups. The first group of 140 incumbents were tested in January 1986 and February 1986. The firefighters first took the hose hoist event which was set up outdoors. Kriska Test. Afterwards, they

proceeded indoors to the stairway climb event and, thereafter, performed the events in the same order as the applicants. However, the apparatus designed to simulate the pike pole broke when used by the first firefighter. Consequently, none of the firefighters in January took the pike pole test. Testing of this group of incumbents also varied from the applicants' testing in that the incumbents were asked to proceed to the next event but were not accompanied by a monitor. Consequently, some firefighters, who were tested by company, waited for other members of their company after the hose hoist and stairway climb events. Approximately one half the firefighters waited after the hose hoist. Id. It is uncertain how many fighters waited after the stairway event, but there is no indication that the numbers were significant.

After the first group of incumbents took the test, the test was modified. The hose hoist event was moved inside and changed in the sequence of events tested. In addition, the pike pole test was re-designed. The test was then administered to applicants. Thereafter, in April and later in May and July, firefighters from the first group returned to re-take the pike pole and hose hoist events.

The plaintiffs claim that the applicants were subjected to a more rigorous test because they were not allowed to wait after any event, except the stairway event, and because they were required to complete all seven events at one time, whereas all but one of the incumbents took the test in two parts. After considering the evidence and the differences in the circumstances under which the incumbents and the applicants took the 1986 capacity test, the Court finds that the above irregularities did not materially

effect the degree of difficulty of the tests taken by the incumbents and applicants.

The evidence indicates that incumbent firefighters primarily waited after the initial hose hoist event. However, the order of this event was changed so that the applicants took the stairway event first. Thus, the fact that incumbents may have rested after the first hose hoist event did not prejudice the applicants since they took no event prior to the stairway climb. Nor does the fact that the firefighters took the hose hoist event later materially effect the applicants since the Court has determined that the hose hoist event must be deleted from the 1986 test. The Court does acknowledge, however, that incumbent firefighters may have had a slight advantage by taking the pike pole event separately instead of at the end of the physical capacity test. This, in turn, might

have resulted in lower comparative scores for a few applicants. The Court does not believe, however, that this advantage was sufficient to warrant re-administration of the entire test to the incumbents. Indeed, considering plaintiffs' own expert's testimony that it is practically impossible to such large scale testing without running into numerous problems, Cranny Test., the Court is extremely reluctant to require the City to re-test the incumbents on the entire test solely because some applicants might have been slightly disadvantaged because of higher scores potentially received by incumbents on the pike pole event.

Part III

In addition to demonstrating that the 1986 physical capability test is content valid, the defendants must also establish a cutoff score "set so as to be

reasonable and consistent with normal expectations of acceptable proficiency within the work force." 29 C.F.R. 1607.14(5)(H). As the Second Circuit stated in Guardians:

> [n]o matter how valid the exam, it is the cutoff score that ultimately determines whether a person passes or fails. A cutoff score unrelated to job performance may well lead to the rejection of applicants who were fully capable of performing the job. When a cutoff score unrelated to job performance produces disparate results, Title VII is violated. See Association Against Discrimination, supra, 594 F.2d at 312-313; Bridgeport Guardians, Inc. v. Bridgeport Civil Service Commission, 482 F.2d 1333, 1338 (2d Cir. 1973). Consequently, there should generally be some independent basis for choosing the cutoff. As with rank-ordering, a criterion-related study is not necessarily required; the employer might establish a valid cut-off by using a professional estimate of the requisite ability levels, or, at the very least, by analyzing the test results to locate a logical "break-point" in the distribution of scores.

Guardians, 630 F.2d at 105.

In the present case, the defendants

propose, as they did for the interim class, a multiple cut score set at one standard deviation below the mean score of the incumbent firefighters for each event. In other words, under the defendants' proposed cut score, an applicant would have to receive a score on each event higher than the score set at one standard deviation below the mean incumbent score for each event. Failure to attain such a score on any event would be deemed a failure of the test as a whole.

In support of their proposed cutoff score, defendants' expert Dr. Jacobs testified that a multiple cut score method was preferable because it was important for firefighter candidates to perform each of the tasks on the physical test because they must perform each event on the actual job. Jacobs Test. He further testified that setting the cutoff level at one standard deviation below the mean was, in his

professional opinion, the appropriate level so as to meet the requirements of section 5(H) of the Uniform Guidelines. Id. Dr. Jacobs also stated that a multiple cut score method was entirely proper because the physical capacity test was highly reliable and had a low degree of inter-correlation. Id. Finally, Dr. Jacobs calculated that over 93% of the incumbent firefighters would pass the 1986 test under the defendants' proposal, thus clearly indicating that the defendants' proposed cut score was "set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force." 29 C.F.R. §. 1607.5(H).

The Court is somewhat disappointed that the defendants have attempted to resuscitate the above proposal which the Court found to be without merit in its Opinion dated October 8, 1986. First, there is absolutely no substance to Dr. Jacobs' estimate that

93% of the incumbents would pass the 1986 test if the defendants' proposed cutoff score was adopted. This figure is based upon the pass rate of an extremely small sample (n = 15) of the total number of firefighters who took the 1986 test for which the defendants have complete data sets (n = 129).[14] Rather, as Dr. Cranny calculated, nearly one-half of the incumbents would have failed the test under the defendants' cut score proposal. See Ex. 5. The implication of such a result is that half the firefighters are unqualified to perform their duties. Such is clearly not the case. Consequently, the only alternative conclusion which the Court can reach is that the defendants' proposed cutoff score proposal is far too high and eliminates individuals who are otherwise fully capable of performing the job. Whereas setting the cutoff score at the level proposed by the defendants would produce a disparate result, see Table 1 to

Exhibit N, the Court concludes that defendants' proposal does not meet the requirement of section 5(H) of the Uniform Guidelines and, therefore, would be violative of Title VII.

The defendants' argument in support of their cut score proposal has further flaws. As Dr. Cranny and Dr. Jacobs testified, a multiple cut score method is only proper where a test has high reliability and the events have low inter-correlations. Jacobs and Cranny Test. However, in the present case, many of the test events measure the same abilities. See Exhibit 5 to Def. Ex. L. This means that there is a high degree of inter-correlation. Therefore, the Court concludes that it would be inappropriate to use a multiple cut score method.

Finally, the Court turns to Dr.Jacobs' rather compelling argument that a multiple cut score method is appropriate because each firefighter should be able to perform each task on the 1986 physical test.

The problem with this argument is directly related to the two points discussed above. To the extent that the 1986 test has events which measure the same abilities, failure of one event does not necessarily mean that a candidate lacks the requisite abilities to be a firefighter. More importantly, the adoption of a multiple cut score system results in significant failure rates among incumbents. Under the defendants' proposal, sixteen percent of incumbents failed a particular event. However, because the same individuals would not fail each event, the cumulative effect of this method would be a failure of forty-eight percent of the incumbents. Ex. 5.

The plaintiffs proposed a cutoff score at a lower level than the defendants. Specifically, plaintiffs propose a compensatory cut score set at the lowest score attained by an incumbent firefighter.

The plaintiffs argue that this is the only valid cut score since the defendants have failed to show that the lowest scoring incumbent is not performing the job of firefighter at an acceptable level.

The Court agrees with the plaintiffs that the City has the burden of showing that a particular cutoff score is valid. Guardians, 630 F.2d at 105. However, the Court finds plaintiffs' proposal to set the cut score at the level of the poorest performing incumbent to be unacceptable. As plaintiffs' own expert testified, every organization includes individuals who are performing at less than proficient levels. Cranny Test. The Court sees no reason to believe that this is any less true for the Columbus Fire Division than any other organization. Consequently, setting the cutoff score at the level of the poorest performing incumbent would result in qualifying applicants as

candidates for the job of firefighter even though they are not so qualified. The Court finds this result to be contrary to section 5(H) of the Uniform Guidelines and, therefore, it must reject plaintiffs' proposal to base a cutoff score on the performance of the lowest-scoring incumbent.

Absent any evidence concerning job performance of incumbents or other credible expert testimony concerning where the cut score should be set, the Court turns to the frequency distribution data submitted by the defendants to determine whether the test results indicate "a logical 'break-point' in the distribution of scores" at which a cutoff score could be set. See Id. at 105; Ex. N, p. 6 and sum of Z's - Incumbent Performance frequency distribution graph. In examining the "sum of Z's" frequency distribution for incumbent performance,[15] the Court observes that there is a natural break-point between the "sum of Z's"

scores of -3 and -5. Approximately sixteen percent of the firefighters scored below this point. The Court also notes that the "sum of Z's" score corresponding to one standard deviation below the incumbent mean is -4.1042, which is obviously midpoint between -3 and -5. Ex. N. Thus, it appears that setting the cut score at one standard deviation below the incumbent "sum of Z's" scores would be appropriate. The Court expressly notes that this is the identical cutoff score used to hire firefighters on an interim basis. The Court, therefore, hereby APPROVES the cutoff score which the City used in conjunction with interim hiring for use on a permanent basis.

Remedy

Whereas the Court concludes that the 1986 physical capacity test is not content valid in its

current form, the Court hereby **ENJOINS** the defendants from using the 1986 physical capacity test as administered by the defendants or results therefrom in hiring future firefighter classes. In addition, the defendants are **ENJOINED** from initiating new interim class until the terms of this Order are fulfilled.

Although the 1986 physical capacity test in its current form is not content valid, the Court concludes that the 1986 physical capacity test without the hose hoist event is content valid. Therefore, the Court hereby **APPROVES** the use of the 1986 physical capability test as so modified and the results therefrom for selecting future firefighter classes. However, because the hose hoist event must be deleted, the defendants must recompute the Z scores of the incumbents as well as the applicants prior to proceeding with any future firefighter classes. In doing so, the defendants should follow the

methodology and procedures outlined by the Court in earlier orders. More specifically, the defendants are ORDERED to do the following:

1) After recomputing the Z scores for incumbents and applicants, determine the number of male and female applicants who passed the 1986 physical capacity examination based upon the cutoff score set at one standard deviation below the mean incumbent sum of Z's score. The sum of Z scores should be computed in accordance with the procedures utilized in constituting the interim classes.

2) Utilize the same cutoff score for the written test that was used in conjunction with the interim firefighters class. Applicants must pass both the written and physical test in order to qualify as a candidate. October 8, 1986 Opinion at 8.

3) Recalculate the scores of the 1984 female applicants who took the 1986 physical capacity test in August 1986.

4) Recalculate the proportions of 1986 male and female applicants who passed the 1986 firefighter test. The defendants must continue to hire male and female firefighters in proportion to their numbers in the candidate population. In calculating the number of females that should be hired, if the number arrived at is a fraction, the number should be rounded up to the nearest whole number.[16]

5) After completing step 4, if a larger proportion of females passed the 1986 physical capability test approved by the Court than the 1986 test with the hose hoist, the defendants must determine whether more female applicants from 1986 and 1984 class should have been included in the first interim class. If so, the appropriate number of spaces should be reserved in the next interim class. In conducting this analysis, the defendants should follow the procedures described in the Court's prior Orders.

6) Once the defendants determine the number of female and male candidates that must be included in each class, it may select the individual male and female candidates by whatever method it desires so long as it is not inconsistent with any prior Orders of this Court.

The Court ORDERS the defendants' to submit a report on its compliance with the above steps by 12:00 p.m. on Thursday, May 28, 1987.

IT IS SO ORDERED.

UNITED STATES DISTRICT JUDGE

[1] The Court specifically reserves judgment as to whether these deficiencies render the scores unreliable for use in a criterion-related validity study.

2
The Ingram job analysis was not introduced as an Exhibit in the second hearing, but was introduced in Brunet I as Jt. Ex. 18. and thus, is part of the record in this case. The Court will hereinafter refer to it as "Ingram Job analysis."

3
This figure is important because it represents the number of female applicants who received scores higher than the cutoff score proposed by the City. It also roughly corresponds with the cutoff score approved by the Court. See infra at 36–41.

4
In the Landy–Jacobs job analysis, the tasks rescue and search were the second and third highest ranked tasks. Appendix F to Ex. K. In the Ingram job analysis, the ability to function without sight, the ability to crawl on hands and knees, and the ability to move around objects weakened by fire are rated very important abilities. The ability to carry or drag adults and/or children was also important, but less so. Ingram Job Analysis Appendix 5.

5
The Ingram job analysis also indicated that the ability to drag empty hose lines, the ability to advance or assist in advancing changed hose lines, and the ability to hoist or lower victims using ropes were relatively unimportant.

6
Other taxonomic abilities which incorporated test events into their definitions included static strength ("drag a victim from the emergency scene or when

the fighter must drag a hose."), stamina ("running a long distance or repeatedly climbing up and down a staircase in full gear."), and extent flexibility ("reaching from a ladder to hang a fan.").

7
With respect to the taxonomic abilities listed in footnote 5, only six out of eighteen psychologists rated static strength above twenty percent on the rescue event. Over half ranked the static strength component at or below fifteen percent for this event. Similarly, only three psychologists ranked event flexibility above twenty percent on the fan hang event.

8
The problem of inflated correlation coefficients resulting from the inclusion of irrelevant abilities in the data set is discussed infra at 22-23.

9
The psychologists' rating of the test overall was as follows (top nine abilities only): Stamina (11.26); Dynamic Strength (7.91); Static Strength (5.38); Dynamic Flexibility (5.15); Speed of Limb Movement (4.50); Explosive Strength (3.81); Multilimb Coordination (3.30); Gross Body Coordination (3.15). The correlation coefficient was .78.

10
The Court notes that these cases discussed the significance of correlation coefficients in the context of criterion-related validity studies. However, they nonetheless indicate that smaller degrees of correlation are acceptable in validation studies.

11
In order to arrive at this table, the Court averaged the psychologists' ratings of abilities across the six test events, excluding the hose hoist event. The bottom seven abilities measured by the test included dynamic flexibility (3.34), manual dexterity (3.03), extent flexibility (1.48), reaction time (.85), gross body equilibrium (.82), choice reaction time (.85), and finger dexterity (.27).

12
The only abilities with ratings which deviate farther from the firefighters' ratings as a result of deleting the hose hoist event from the test are gross body equilibrium, wrist finger speed, gross body coordination and manual dexterity.

13
The pike pole event also significantly impacted upon women. As discussed on page 11, supra, only two female applicants completed more repetitions than the average firefighter and a total of six female applicants did better than the bottom fifteen percent of incumbents. However, the pike pole event is rated as a fairly important event by firefighters. Thus, the fact that it has a disparate impact upon women does not run afoul of the requirements of Title VII.

14
For reasons that have never been satisfactorily articulated to the Court, Dr. Jacobs began his analysis with a "norming group" comprised of 54 incumbents, instead of the 129 for which a full set of data was available. He then reduced this group from 54 to 40 by eliminating statistical outliers which he categorized as

those individuals with the five percent best and worst scores. Not only is such a simplistic method of eliminating outliers statistically incorrect, see J. L. Gill, Design and Analysis of Experiments, 153-154 (1978), even if ten percent of the scores were outliers, Dr. Jacobs figures do not add up, e.g., 54 - .1(54) = 49. Of this norming group, only 67% of the incumbents would pass - an unacceptably low pass rate under the Guidelines since it would mean that 33% of the incumbents were unqualified. Dr. Jacobs then looked at a subset of this subset of a subset of the population to determine that 93% of the firefighters would pass the test. Such a statistic is wholly unacceptable.

15

A "sum of Z's" score is the scores of an incumbent fighter expressed in terms of standard deviations from the mean on each event summed over all events. In this case, the square root of standard deviations exceeding one standard deviation above the mean were taken in order to reduce the emphasis on speed.

16

The Court seeks to avoid regular disputes concerning whether a fraction resulting from proportional hiring, e.g., 2.3 or 2.7, must be rounded up or down and whether the City deliberately made a smaller class in order to avoid hiring an extra female candidate. Consequently, the Court believes that a simple rule of thumb should be applied in order to conserve the Court's and the parties' time and resources.

UNITED STATES DISTRICT COURT
SOUTHERN OHIO DISTRICT
EASTERN DIVISION

ANN BRUNET, et al.,
Plaintiffs,

v. Case No. C-2-84-1973

CITY OF COLUMBUS, et al.,
Defendants.

OPINION AND ORDER

In accordance with the terms of this Court's Order of May 21, 1987, the defendants submitted a report regarding their compliance with the procedures outlined in that Order. Plaintiffs filed an appropriate response to defendants' report. In addition, on June 5, 1987, defendants filed a rather unusual motion, captioned a "motion for consideration of defendants' rights in the remedy phase." At the same time that they filed this extraordinary motion, defendants filed a motion for expedited consideration of the motion,

on the grounds that any delay in its consideration would further delay the initiation of a new interim class and cause financial loss to the defendants. The pending motions and the parties' calculation reports will be addressed in this Order.

Having carefully reviewed the parties' calculations, the Court concludes that the male/female pass ratio is .03189. In accordance with the standard mathematics practices, this number should be rounded up to 3.19 percent. As the plaintiffs correctly point out, applying this percentage against the number of 1984 female applicants taking the test results in a figure of 4.019, or 4.02. Pursuant to the Court's Opinion dated May 21, 1987, the plaintiffs argue that this number should be rounded up to the nearest whole number and, consequently, that a total of five (5) female applicants should be hired from the 1984 class.

Although the plaintiffs correctly apply the rule of thumb adopted by the Court for resolving "rounding" problems, the Court only intended that this rule be applied to the 1986 applicants since no harm would result from rounding up fractions. However, application of this rule of thumb in the calculation of the number of 1984 female applicants which should have been hired could harm the defendants since they face greater backpay liability if more female applicants are hired. The Court believes that it would be manifestly unfair to impose backpay liability on the City solely because a tiny fraction was rounded up for administrative convenience. Moreover, the concerns which prompted the Court to adopt the above rule of thumb, namely the manipulation of class size to reduce any fraction to below .5 and disputes inherent in such controversies, do not apply with respect to the 1984 firefighter class since the class size is fixed

(126 class members) and this issue will be non-recurring.

For these reasons, the Court believes that it should apply standard mathematical practices and round the 4.02 figure down to 4.00. Consequently, the Court finds that a total of four female applicants should have been hired from the 1984 class. Given that the defendants have already hired three of the female applicants from the 1984 class, the Court hereby ORDERS the defendants to hire an additional female applicant from the 1984 class for the next firefighter class. Thereafter, the defendants need not hire any additional 1984 female applicants.

Whereas the defendants have satisfied the terms of the Court's Order of May 21, 1987 with respect to the necessary calculations, the injunction against the initiation of new interim classes is hereby VACATED.

However, with respect to all future classes, the Court believes that the parties should jointly determine the appropriate number of women and men to be hired from the applicant class. These joint calculations should than be submitted to the Court for its review. Therefore, the Court DIRECTS the parties to submit agreed calculations specifying the number of men and women to be hired in all future classes.

Finally, in their artfully captioned "motion for consideration of defendants' rights in the remedy phase," the defendants argue that if the 1984 physical capacity test is regarded merely as a qualifying examination, none of the 1984 females would have been hired based on the scores they received on the cognitive abilities test. The Court believes that the issue raised by the defendants falls within the scope of its Orders of May 13 and May 31, 1986, both of which are currently pending on appeal. Consequently,

the Court does not have jurisdiction to consider defendants' motion. The motion is, therefore, DENIED without further consideration.

IT IS SO ORDERED

_____

UNITED STATES DISTRICT JUDGE